Michael Rea

Metaphysics

the basics

Second Edition

METAPHYSICS

THE BASICS

Metaphysics: The Basics is a concise and engaging introduction to the philosophical study of some of the most important and foundational aspects of the world in which we live. Concerned with questions about existence, time, identity, change, and other basic elements of our common-sense and scientific ways of thinking about the world, metaphysics has long fascinated people. But to the uninitiated, many of the issues and problems can appear bewilderingly complex and intractable. In this lively and lucid book, Michael Rea examines and explains the core questions in the study of metaphysics—questions such as:

- What is the relationship between an object and its properties, or between an object and its parts?
- What is time, and is time travel possible?
- Are human beings free?
- What is it for an object or person to persist over time?

This second edition has been thoroughly revised and includes a new chapter on the metaphysics of gender. With suggestions for further reading and a glossary of key terms, *Metaphysics: The Basics* is an ideal introduction for those coming to the subject for the first time.

Michael Rea is Rev. John A. O'Brien Professor of Philosophy at the University of Notre Dame, USA, and Professorial Fellow in the Logos Institute for Analytic and Exegetical Theology at the University of St Andrews, UK. He has written or edited more than fifteen books, and has given numerous lectures around the world, including the 2017 Gifford Lectures at the University of St Andrews.

THE BASICS

The Basics is a highly successful series of accessible guidebooks which provide an overview of the fundamental principles of a subject area in a jargon-free and undaunting format.

Intended for students approaching a subject for the first time, the books both introduce the essentials of a subject and provide an ideal springboard for further study. With over 50 titles spanning subjects from Artificial Intelligence to Women's Studies, *The Basics* are an ideal starting point for students seeking to understand a subject area.

Each text comes with recommendations for further study and gradually introduces the complexities and nuances within a subject.

AMERICAN PHILOSOPHY
NANCY STANLICK

ANIMAL ETHICS
TONY MILLIGAN

ARTIFICIAL INTELLIGENCE
KEVIN WARWICK

EVOLUTION
SHERRIE LYONS

FOOD ETHICS
RONALD SANDLER

FREE WILL
MEGHAN GRIFFITH

METAPHYSICS
MICHAEL REA

PHILOSOPHY (FIFTH EDITION)
NIGEL WARBURTON

GLOBAL JUSTICE
CARL DEATH AND HUW L. WILLIAMS

HUMAN GENETICS (SECOND EDITION)
RICKI LEWIS

LOGIC (SECOND EDITION)
JC BEALL AND SHAY ALLEN LOGAN

BIOETHICS (SECOND EDITION)
ALASTAIR V. CAMPBELL

EASTERN PHILOSOPHY (SECOND EDITION)
VICTORIA HARRISON

PHENOMENOLOGY
DAN ZAHAVI

ATHEISM
GRAHAM OPPY

EMOTION
MICHAEL BRADY

PHILOSOPHY OF MIND
AMY KIND

METAPHYSICS (SECOND EDITION)
MICHAEL REA

Other titles in the series can be found at: www.routledge.com/The-Basics/book-series/B

METAPHYSICS

THE BASICS

Second Edition

Michael Rea

For Gretchen

CONTENTS

PREFACE TO THE FIRST EDITION

Western philosophy began with metaphysics. The earliest Greek philosophers were on a quest for the underlying natures of things. Some said everything is ultimately water; others said everything is fire; still others said that it is air. Pythagoras thought that everything was ultimately made up of mathematical objects (and *that* mystical view still has adherents even today). They also worried about change. The world is all in flux, observed Heraclitus—so much so that it is impossible to step into the same river twice. But what does *that* mean? You can step into the Nile twice; you can step into the Mississippi twice. So if it is really impossible to step into the same river twice, does that mean that there is something illusory about the Nile and the Mississippi, and perhaps about everything else that is changing? Parmenides said yes. Many in his wake decided that ultimate reality had to consist only of *unchanging* things. No more water, fire, and air; now it was ungenerated and indestructible atoms, or abstract ideas, or other esoteric things. And on and on.

Metaphysics in the 21st century deals with these questions and much, much more. It is a core area of philosophy and has flourished as such throughout most of the history of the discipline. And with good reason: metaphysics deals with some of the deepest, most existentially important questions about human persons and the fundamental

features of reality. Many of the questions that matter most to us—can we survive death? are we free? what is it to be a person? do we have immaterial souls?—depend on decisions about matters metaphysical. Even the apparently esoteric answers described in the first paragraph to questions about change and the ultimate nature of things have proven important, exerting tremendous influence not only on the history of philosophy but on the history of theology as well, impacting even our very concept of God.

This book aims to provide an accessible introduction to the basics of metaphysics. Metaphysics is an abstract subject. Rarely does it make for breezy reading. But I have tried as much as possible to keep things simple, to avoid the jargon of the trade and to expose the structure of the arguments under discussion so that readers can see with clarity what is going on. In some cases, the inclusion of technical terms was unavoidable. I have bolded the first substantive occurrences of terms like this that recur multiple times in the text, and I have supplied definitions or explanations in the Glossary at the end. The Glossary does not add much beyond the definitions I supply in the text itself; but its purpose is to allow the reader to quickly access those definitions without having to look for them in the various chapters in which they are given.

The book begins with a chapter that explains what metaphysics is and what some of the main criticisms of metaphysics have been. Subsequent chapters focus on six topics that have been central to the discipline both throughout its history and in the contemporary literature. My hope is that this text will be of use not only to people who are looking to be introduced to some of the main problems and perspectives that have taken center stage in the contemporary literature in metaphysics, but also to people who just want to reflect on some of the great problems that have occupied the minds of great metaphysicians throughout the history of philosophy.

My approach, then, is *problem-based* rather than *figure-based*: I focus more on the problems that have occupied metaphysicians rather than on the metaphysicians themselves. But I have tried throughout to provide references both within the text and in the "Further Reading" that will guide the reader to some of the most important defenders (both contemporary and historical) of the perspectives developed herein. Also, I should note that, following the general format for this

series, I have not provided in-text citations for the philosophers I discuss or the texts from which I quote throughout the book. Rather, in each chapter I have simply indicated whom I am discussing or quoting and, where necessary, I have also given titles; then I have listed the works discussed or quoted from in the Further Reading section at the end. I have written this book so that a natural companion for it would be my edited textbook, *Arguing about Metaphysics* (also published by Routledge). The two books treat the same topics (except that the present volume includes a chapter on "substance" whereas *Arguing about Metaphysics* does not have a special section on that topic), and many of the specific articles discussed in this book are reprinted in *Arguing about Metaphysics*.

I am grateful to several friends and colleagues who offered valuable help at various stages in the process of writing this book. The discussions of temporal experience and time travel in Chapter 3 were heavily influenced by discussions of papers written by me and by L. A. Paul in the Notre Dame Metaphysics & Philosophy of Religion Reading Group. Contributions to those discussions by Meghan Sullivan and Kenny Boyce were particularly helpful. Andrew Bailey provided me with very helpful comments on early drafts of the first three chapters, and Siobhán Poole reviewed the manuscript for the press and offered many suggestions for making things shorter and more accessible. For all of this I am very thankful. I am especially grateful to Alicia Finch and Jeff Snapper, both of whom commented on all of the chapters. Whatever its remaining failings might be, the book is much clearer and more accessible now thanks to their generously extensive and insightful comments. Finally, I am most grateful to my wife Chris and to my children, Aaron, Kristina, Gretchen, and Matthias, all of whom graciously put up with me during those periods when the writing of this book dominated far too much of my work and free time.

PREFACE TO THE SECOND EDITION

The present edition of *Metaphysics: The Basics* offers a significant reworking of the structure of the book, with an eye both to making it still more user-friendly and to enable instructors to use just portions of the book in classes without assigning the whole text. The previous edition had six rather long chapters, whereas the present edition divides (mostly) the same material into eleven chapters and adds one additional chapter of new material.

The new chapter focuses on the metaphysics of gender and, because issues about "social construction" play an important role in the literature on social metaphysics, it also includes substantial discussion of mind dependence and related notions. As I note in that chapter, gender is just one among several important topics in social metaphysics that have been the subject of a great deal of recent interest. I would have liked to have been able to survey a broad range of those topics; but, in the end, I decided that treating one such topic in greater depth would be of more use to readers than a relatively shallow survey of a broader range of topics. Gender is one of the topics that has received the most attention in the years since the first edition of this book was published, and it has been the focus of much of my own recent research and teaching.

I will not repeat here the acknowledgments that appear in the original Preface to this book, but I remain deeply grateful to the friends, colleagues, and family members mentioned there for their help and support. My daughter Penelope, who was not yet born when I wrote the first edition, ought now to be added to the list of supportive family members. I would also like to add my thanks to the anonymous reviewers for Routledge who provided feedback on the first edition that informed the revisions for this second edition of the book, and to Jc Beall and Daniel Nolan for advice on some technical details in a couple of the chapters; and I am especially grateful to Elizabeth Barnes, Sara Bernstein, Lindsey Breitweiser, Laura Callahan, and Robin Dembroff for their helpful comments and advice on earlier drafts of the chapter on gender.

INTRODUCTION

Most of us have questions about the world that science does not answer. Some of these are questions about what exists, questions about the intrinsic natures of things, questions about how things in the world have to be or about how they could have been different, or questions about the fundamental structure of the world. These are the sorts of questions with which this book is concerned.

For example: My words here are causing you to have certain thoughts. But what is the *nature* of the causal relation? What is it, exactly, for one thing to cause another? Is causation some sort of necessary connection between objects or events? When a cause occurs does its effect *have* to follow afterward? If you are caused to do something, does that mean that you do not do it freely? For that matter, do you do anything freely, or is everything you do determined by the laws of nature? What does it even mean to say that you are free, or that there are such things as laws of nature? Is there a God? Do you have an immaterial soul? When you say "the number three", is there some actual thing that you are referring to that is prime, odd, and somehow in between two other things, namely the numbers two and four?

For some people, questions in metaphysics are conversation stoppers. Ask, over coffee, what it means to be free and you might be answered with a roll of the eyes, or a profound, tight-lipped, "meaningful nod". For others, however, they are scintillating puzzles—invitations to

unravel a myriad conceptual knots in our commonsense and scientific ways of thinking about the world.

The eye-rollers tend to convey the impression that they have never even thought about metaphysical questions. The distributors of meaningful nods indicate by their nodding that *of course* we all know that metaphysical questions are profound but unanswerable. But in fact most people have at some point in their lives both thought about metaphysical questions and thought that those questions could be answered. Many of us begin reflecting on matters metaphysical in early childhood. We wonder what it could mean to say that God is everywhere. We wonder about the passage of time: If time moves, what does it move *in*? (Not in time, of course. But then what?) We wonder about ourselves: Do we control our thoughts? Can we exist outside our bodies? And so on. To the extent that we entertain and try to answer questions like this, we are *doing* metaphysics. But what exactly *is* metaphysics? My goal in this chapter is to provide an answer to this question.

Bookstores often have entire sections devoted to "metaphysics". They tend to be filled with books that deal with occult topics like astrology, ghosts, psychic powers, the secret lives of plants, and the like. These issues are what folks in the general public recognize as falling under the label "metaphysics". So if you tell your dental hygienist or your prospective in-laws that you are studying metaphysics in college they are all too likely to think that you are devoting your time and attention to something strange and frivolous rather than to a serious academic subject. But in fact these topics have very little to do with those studied by academic metaphysicians.

So what *can* we metaphysicians and students of metaphysics tell our friends and family about the subject to which we are devoting so much of our time? Here are a few common answers, followed by the one that I myself prefer.

BEING AS SUCH

Aristotle famously characterized metaphysics as the study of *being qua being*, or of *being as such*. A simpler way of putting the same idea is to say that metaphysics, according to Aristotle, is an investigation into the

different kinds or categories of being. To understand what he had in mind, it is helpful to know that, as Aristotle saw things, terms like "being" or "existence" have a variety of different meanings. For example, what it is for a horse to exist is very different from what it is for a number to exist. You might think that this is just a funny way of expressing the commonsensical idea that a horse is a different kind of thing from a number. But that wouldn't be quite right. A horse is a different kind of thing from a cat; but what *existence* is for a horse is the same as what existence is for a cat. They have the same kind of being, even if they are not exactly the same kind of object. Numbers, on the other hand, besides being different kinds of objects from horses or cats, do not even *exist* in the same way. They have a very different kind of being.

Aristotle's idea was that "being" and related words (like "is" or "exists") mean one thing when predicated of a horse or a cat and something else when predicated of a number. The meanings are related to one another, but still different. We can understand this point by way of an analogy. Consider the word "healthy". We might say that a meal is healthy; we might also say that someone's complexion is healthy, or that they themselves are healthy. It seems that "healthy" means something different in each of these three cases. The meanings are related; but they are still different. So likewise, Aristotle thought, with words like "exists" or "is": they too vary in meaning, depending on the sort of thing to which they are applied.

According to Aristotle, then, the fundamental task of metaphysics is to discover and more richly understand the most general kinds or categories of being. In carrying out this task, what we are most interested in are the different meanings of the term "being", which correspond to the different categories. We are not so interested in what it is *to be a horse*, or *to be the number three*. Rather, we are interested in what it is *to be,* in each of its different senses. In other words, we are interested in being *as such*, and not in *beings* themselves and their particular distinguishing attributes.

The trouble with this characterization is that most philosophers nowadays think that there is a lot more to metaphysics than the study of being as such. Questions about the compatibility of freedom and **determinism**, or about the nature of the causal relation, for example, don't seem to be questions about being as such, but they are generally regarded as belonging to the domain of metaphysics. Likewise,

non-scientific questions about what there is—whether there are abstract objects, for example, or a divine being, and questions about what such things would have to be like if they did exist—are generally regarded as metaphysical questions; but it is hard to see them as questions about being as such. Moreover, the view that these kinds of questions are metaphysical questions isn't a recent development. It is how the field has been conceived for hundreds of years.

ULTIMATE REALITY

According to a more common characterization, metaphysics is the study of *what there is*, or of *what there REALLY is*, or of *ultimate reality*. But these characterizations too are less than helpful. Why think that it is in metaphysics, rather than in botany or zoology or theoretical physics, that we learn about "ultimate reality" or about "what there really is"? Why are we doing metaphysics when we ask whether there are numbers or sets, but not when we ask whether there are unicorns? Why do questions about the nature of causation have more to do with ultimate reality than do questions about the function of a human heart, or about the defining characteristics of electrons? What does it even mean to make a distinction between what exists and what *really* exists, or between reality and *ultimate* reality?

We can get some insight into the idea underlying this characterization when we look at the difference between the way in which metaphysicians ask certain kinds of questions and the way in which scientists, mathematicians, or people in the ordinary business of life ask those same questions. Is there a table in this room? As an everyday sort of question, the answer is settled by a quick look around. We take for granted the idea that, as a general rule, table-experiences are caused by tables. So, after looking around, if we find ourselves with table-experiences we say, "Yes, there is a table in this room".

If we are doing metaphysics, however, part of what we are calling into question is the assumption that our table-experiences are caused by (or are in some other way dependent upon) objects answering to the general description of a *table*. We don't normally believe that *football teams* are large objects made up of players. We normally think that our football-team experiences are caused not by a single object but by a bunch of objects acting in concert. Why not say

the same thing about (so-called) tables? In other words, when you have a table-experience, why suppose that there is a single object—a table—causing that experience? Why not suppose instead that *there is no table*, but just a bunch of atoms arranged table-wise?

We can now begin to see why people say that metaphysics is concerned with questions about what *really* exists, or about what is *ultimately* real. "Is there a table in this room?" is naturally construed as an ordinary question with a straightforward answer. But if we ask, "Is there *really* a table in this room?" or "Are tables part of *ultimate* reality?" we signal our interest in some of the further, metaphysical questions raised at the end of the previous paragraph. Depending on how we answer these questions, we might conclude that the "straightforward" answer is false, even if it is perfectly appropriate for the ordinary business of life. Or, depending on our views about language, we might conclude that it is true, but somehow not in conflict with the answer we give at the end of our metaphysical inquiry—just as we take there to be no conflict between the "ordinary" claim that the sun is setting over the mountain and the more scientifically informed claim that the Earth's rotation is causing the mountain to obscure our vision of the sun. Either way, the main point to note is that the questions we are interested in when we ask whether there is *really* a table in the room are not answered by science or observation alone. Instead, the answers depend upon the truth or falsity of more general metaphysical claims about what it takes to make several objects become parts of a larger whole, or about what it is to be a material, and so on.

Still, as a general characterization of metaphysics, the idea that metaphysics is the *study of what really exists* falls short because it is incomplete. The question whether free will is compatible with determinism, for example, is a paradigmatic example of a question in metaphysics. But it is not naturally construed as a question about what really exists, or even as a question about ultimate reality.

THE FUNDAMENTAL LEVEL

For the past couple of decades, there has been a growing trend among metaphysicians of saying that metaphysics is (or should be thought of as) the study of what is **fundamental**, or *basic*. This way of thinking

about metaphysics has also been attributed to Aristotle, and it has been gaining traction since the publication of Jonathan Schaffer's influential paper, "On What Grounds What". Schaffer's paper opens by declaring his intention to revive a conception of metaphysics that differs from the one that he then took to be dominant in the field. On the dominant conception, he says, "metaphysics is about what there is. Metaphysics so conceived is concerned with such questions as whether properties exist, whether meanings exist, and whether numbers exist". On the conception he aims to revive, metaphysics "does not bother asking whether properties, meanings, and numbers exist. Of course they do! The question is whether or not they are *fundamental*".

Fundamental things, if there are any, are the things out of which all others are made, or the things without which nothing else could even exist. They are the "grounding" entities, the things on which everything else in the world depends. For example, Thales famously said that everything is water. The view seems absolutely incredible if we think that what he meant was that, contrary to all appearances, there's no real difference between a baseball and a raindrop. But the standard way of understanding Thales' famous claim is as a claim about what is fundamental. The underlying stuff in the world is water; everything is ultimately made out of water; water is the fundamental stuff of the universe. Similarly, many people nowadays endorse **physicalism**, which can be roughly characterized as the view that everything is ultimately made up of properties and objects posited by our best theories in physics. This, too, is naturally understood as a thesis about what is fundamental.

The view that metaphysics is concerned with what is fundamental is obviously related to the idea that metaphysics is concerned with what "really" or "ultimately" exists. "Ultimate" is sometimes used as a near synonym for "fundamental", so it is easy to see why the idea that metaphysics is concerned with ultimate reality might be understood as the idea that metaphysics aims to discover what is fundamental. Similarly, a metaphysician who says, for example, that table-experiences are caused not by tables but simply by atoms arranged table-wise might express their view by saying "there aren't *really* any tables". But since the very same metaphysician will probably, in the ordinary business of life, do things like ask their children to come sit at the table for dinner, we might think that all they mean by saying

that there aren't really any tables is that tables are not fundamental, that they are instead reducible to atoms or particles.

As with previous characterizations, however, the problem with this one is incompleteness. Metaphysics *is not* just about what is fundamental. Historically, the following questions have been almost universally acknowledged as falling within the domain of metaphysics:

- Is change really possible? If so, what does it mean to say that something has changed?
- Can the past be changed? How about the future?
- Is the passage of time possible? What is time, anyway?
- What is an event? Can the same event happen more than once? What is involved in one event's *causing* another?
- What are human minds? Are they immaterial thinking substances, or are they material objects (brains, perhaps?), or something else entirely?
- Are there any non-physical things? If so, could they causally interact with physical things?
- Are human beings free? Is freedom even possible?
- Is it possible to live after death?
- Do human beings or human faculties have anything like a proper function?

None of these is plausibly construed as a question about what is fundamental.

I suspect that claims like "metaphysics is about what is fundamental" are not usually intended to be taken strictly at face value. They are most charitably construed as hyperbolic claims intended to push metaphysicians into a particular way of thinking about their discipline. The goal isn't to get metaphysicians to see questions like those just mentioned as falling outside their discipline, but rather to get them to see questions about fundamentality and grounding as in an important way *central* to their discipline.

Even so, I must still register skepticism. Why think that questions about what is fundamental are in any way more "central" or "important" than questions about human free agency, or the possibility of change? So far as I am aware, there is no argument to be had for the conclusion that they are. So it is hard to see why we should think anything more than that questions about fundamentality are *among* the

questions of metaphysics—which, of course, no one would ever have denied.

THE FINAL CHARACTERIZATION

We do best to approach the question of what metaphysics is by first looking at what sorts of questions metaphysicians typically ask, and then asking what, if anything, those questions share in common. We have already seen quite a few sample questions; here are a few more:

- When you assert something is there some *thing* that you're asserting? If so, what kind of thing is it?
- When you talk about ways things could have been (e.g., there's more than one way this class could turn out), to what sorts of things (if any) does the word "ways" refer?
- What is the relation between an object and its properties? Are the properties of a thing parts of it? If so, then do objects have any other kinds of things as parts—say, a bare substratum that *has* all of the properties? If not, then are we supposed to imagine that properties are somehow external to the things that have them, and are related to those things simply by resemblance or some other sort of relation?
- A gold statue is constituted by a lump of gold. The lump of gold still exists after it is melted down and reshaped, but the statue doesn't. So, what is the relationship between the lump and the statue? Are they two different things in the same place at the same time? If not, then how do "they" have different survival conditions?
- Are there contingent beings, or is the way things are somehow the way things have to be? Is there a necessary being (a God, perhaps) who created all contingent things?

There are many more such questions, but these are sufficient to provide a feel for the sorts of issues with which metaphysicians typically concern themselves.

What do these questions have in common? I used to say that what they have in common is that they are non-scientific questions about what exists, and about the necessary connections among certain kinds of concepts—namely, those falling outside the domains of ethics, epistemology, logic, and socio-political theory. Although I still think that this is reasonably accurate, putting it this way suggests that there is no methodological continuity whatsoever between

metaphysics and science. It also suggests that metaphysics is a sort of hodge-podge discipline, encompassing all and only questions that haven't been taken over by another sub-discipline. As my colleague Meghan Sullivan put it to me in conversation, this is a "garbage bag" conception of metaphysics; and, more importantly, it does not really capture what it is that *unifies* the discipline of metaphysics. It would be nice if we could do better.

In the previous edition of this textbook, I said that questions in metaphysics have the following in common: To answer them, one must make non-**empirical** claims (i.e., claims that are not based on observation or experience) about what there is or could be, or about the nature or defining essence of some concrete thing, or about the proper analysis of concepts of a certain kind. What kind? At the time I wrote the first edition, I thought the answer was "Just about any kind *other than* those used specifically for the evaluation of the actions, beliefs, or reasoning processes of agents and the institutions created by agents". So, as I thought of it then, most moral and aesthetic concepts, a variety of social, psychological, legal, and political concepts, the concept of knowledge (which, under some analyses, includes the concept of "justification"), and many others would all be excluded from the purview of metaphysics. I excluded these latter concepts because I took their analysis to be central to other fields of philosophy—most notably, ethics, socio-political philosophy, epistemology, and aesthetics. I now think, however, that excluding these concepts from the purview of metaphysics is a mistake. This is not to say, of course, that it is a mistake to think that the analysis of these concepts is indeed central to other fields. Rather, the mistake lies in thinking that metaphysics does not *overlap* with those other fields in dealing with the analysis of their central concepts.

My final characterization of what questions in metaphysics have in common, then, is this:

> Questions in metaphysics are non-empirical questions about what exists (in every sense of the term "exists"), about the nature or defining essence of some thing (in the broadest sense of the term "thing"), or about how things must be or how they could have been different.

Note that I have left out of this characterization reference to the "analysis of concepts", not because I think that metaphysicians aren't concerned with analyzing concepts (far from it!), but rather because I think that analyzing a concept (whether it be an abstract concept like *freedom* or *love*, or a social concept like *money* or *politeness*, or a concept of some object-kind like *human being, horse,* or *table*) is just one way of inquiring into the nature or defining essence of something. Note, too, that I am here using the term "non-empirical" somewhat imprecisely to mean what would normally be expressed by "not entirely empirical" or "not purely empirical". Questions in metaphysics, and their answers, and metaphysical theorizing often take empirical data into account. Moreover, one might reasonably be skeptical, as Quine was, of the idea that there is a sharp distinction to be drawn between the empirical and the non-empirical. But the more precise mode of expression is unwieldy and, because of this, it is also potentially confusing; so I have opted to stick with the term "non-empirical" and to impress upon the reader that the term is not meant to convey that metaphysics is *entirely* free of empirical influence.

The final characterization here adequately captures the fact that (a) metaphysics is a non-empirical mode of inquiry, (b) it is partly about *what there is*, (c) it is partly about describing the essences or natures of things, and (d) it is concerned with what is possible, necessary, or impossible. On this way of thinking about the discipline, metaphysics is not defined solely by its method, nor is it defined solely by its subject matter. Although I have not explicitly identified questions about "being as such" or about "what grounds what" as among those that are investigated by metaphysics, neither does this conception locate them outside the discipline. In fact, they do fall within the domain of metaphysics under this conception; for answering those sorts of questions will involve exactly what I have said that answering metaphysical questions will involve: one will have to make non-empirical claims about the natures of things, about how things can be or must be, and about what there is.

People sometimes distinguish metaphysics from **ontology**, which is often seen as a sub-field within metaphysics. Ontology is commonly characterized as the study of *what there is*. Or, more precisely: it is commonly seen as the branch of study which focuses on *existence* claims of the sort studied by metaphysics, and on the logical consequences

thereof. So, on this way of thinking, metaphysical claims that tell us what exists or that tell us about the natures of things would generally belong to the domain of ontology, whereas metaphysical claims that simply tell us what there *could* be or about how things *must* be (either absolutely, or given certain background conditions) generally wouldn't. I am not myself committed (or even attracted) to the idea that there is any helpfully precise distinction to be drawn between ontology and metaphysics. but the distinction is widespread enough that one must take note of it in an account of the nature of metaphysics.

FURTHER READING

The opening books of Aristotle's *Categories* and *Metaphysics* provide his own characterization of metaphysics. For a contemporary defense of a broadly Aristotelian conception of metaphysics, see Kathrin Koslicki's "Metaphysics: The Science of Essence", in Javier Cumpa (ed.), *The Question of Ontology* (Oxford: Oxford University Press). A good single source for discussion and display of the other conceptions of metaphysics discussed in this chapter is David Chalmers, David Manley, and Ryan Wasserman (eds), *Metametaphysics: New Essays in the Foundations of Ontology* (Oxford: Oxford University Press, 2009), hereafter cited simply as *Metametaphysics*.

   I mentioned several philosophers who take metaphysics to be the study of what is fundamental. A good starting place for this view is Jonathan Schaffer, "On What Grounds What", in *Metametaphysics*, 347–383. Interested readers will also want to see L. A. Paul, "Building the World from Its Fundamental Constituents", *Philosophical Studies* 158, no. 2 (2012): 221–256 and Theodore Sider, *Writing the Book of the World* (New York: Clarendon Press, 2011). As we saw, Schaffer denies that metaphysics is most centrally about "what there is"; but, as we also saw, some think that questions about what there is are best construed as questions about what is fundamental. On this, see Ross P. Cameron, "Truthmakers and Ontological Commitment: Or How to Deal with Complex Objects and Mathematical Ontology Without Getting into Trouble", *Philosophical Studies* 140, no. 1 (2008): 1–18; and "How to Have a Radically Minimal Ontology", *Philosophical Studies* 151, no. 2 (2010): 249–264.

2

THINGS THAT DON'T EXIST

The ontological question, according to Quine, asks "What is there?" Biologists, chemists, physicists, zoologists, and many others supply us with some provisional answers: There are trees, cells, bosons, horses, and so on. Reasoning about our commonsense beliefs can supply us with further answers. Ontologists look more deeply into such provisional answers and ask which of them are true and which only appear to be true.

One might wonder why we should *want* to do ontology. It is annoying to respond to a question like "Where are my keys?" by saying "Well, how do you even know you have keys? Do keys *really* exist?" Commonsense belief in the reality of keys is *useful*. It helps get us around in the world. It is "empirically adequate": It is not the sort of belief that will be seriously called into question by observation or experiment. But if empirically adequate beliefs about what exists are all that we need to get around in the world, why should we try to look any deeper?

The answer is that sometimes ontological questions do bear on matters of great concern. Although it is hard to see how questions about the reality of keys make a difference to anything we care about, it is not at all hard to see the importance of questions like "Is there a

God?" or "Do I have a soul that can survive the death of my body?" or "Am I free" or "Do my commonsense **intuitions** about what exists in the world and about what the world is like lead me badly astray?" (Intuitions are intellectual experiences of the sheer obviousness or necessary truth of various claims. Commonsense intuitions are experiences like this that are widespread and generally thought to be part of our shared stock of common knowledge.) Our answers to these questions will be intimately connected with various ontological questions. Sometimes ontological questions *are* the questions we care about. Sometimes they bear only indirectly on the questions we care about. Either way, they are still important.

Suppose you think that there are immaterial souls and that they are invisible and not physically located in spacetime. Someone might tell you that it is unreasonable to believe in things like that. Can you sensibly reply that it is reasonable to believe in numbers, and say that numbers are like that? That depends on whether it is reasonable to believe that there *really are* numbers. Similarly, as we shall see in a later chapter, it is remarkably easy to show that commonsense beliefs about material objects lead us into contradictions. Learning this, one might just shrug it off and concede that our commonsense conceptual framework is an incoherent mess. Many of us, however, would prefer to see whether we can avoid such contradictions by thinking harder and more carefully about what really exists. It is disturbing to think that our commonsense intuitions are wildly unreliable. So many philosophers have been interested in examining those intuitions to see just how much commonsense ontology can be saved. Sometimes the result of this process is that we reach conclusions like "there are no non-living composite objects", which then have implications for questions about whether our keys or our dining room furniture really exist. So even if those latter questions are not, all by themselves, of philosophical interest, they come to be of interest because of their connection to other questions which are of deeper and more lasting philosophical interest.

As a general rule, people avoid believing in things that are an uneasy fit with their background ideas about what the world is like. So, for example, people (and there are many) who think that whatever exists can, in principle, be investigated by the empirical methods of science are generally reluctant to believe in things (like

immaterial souls, or magical creatures like leprechauns and fairies) that seem, by nature, to be deeply resistant to scientific investigation. People (and this is most of us) who think that reality is non-contradictory are generally reluctant to believe in things (like nonexistent beings) whose very nature seems paradoxical. I suspect that these general preferences go a long way toward explaining why, in contemporary ontology, questions about the existence of abstract objects, nonexistent objects, creatures of fiction, God, souls, and composite material objects of various kinds have tended to dominate the literature.

Abstract objects are things like the number two, the set of all horses, beauty, and so on. They are the sorts of things that cannot even in principle be detected by sensory experience or instruments. (We see beautiful things, not beauty itself; we see horses, but not sets thereof; etc.) If they exist at all, they are immaterial, (probably) have no location in spacetime, and lack causal powers. They are weird. Many philosophers do not wish to believe in them. Belief in God and souls has seemed problematic for similar reasons, though God and souls are usually classed as concrete objects by virtue of their alleged causal powers. Creatures of fiction and nonexistent objects are not only weird but paradoxical. For example, we seem often to talk about such things. Peter Parker (a.k.a. Spiderman) works at *The Daily Bugle* and engineered his own web-shooters. But neither Parker nor *The Daily Bugle* nor Spiderman's web-shooters exist. How could *both* of the previous two sentences be true? Ponce de León spent a good deal of his life searching for the Fountain of Youth, which seems to imply that he spent a good deal of time searching for a *thing* that *does not exist*. But it sounds contradictory to say that there is a thing that does not exist. Composite material objects also present us with paradoxes—enough, in fact, that some philosophers have decided that the best overall response is simply to deny the existence of composite objects altogether.

In the present chapter, after making some preliminary distinctions and introducing a bit of terminology, we shall focus primarily on questions about nonexistent objects and creatures of fiction. In the next chapter, we examine questions about the existence of (certain kinds of) abstract objects.

PRELIMINARIES

As we have already seen, it is helpful to think of ontology as examining what we say loosely speaking and in the ordinary business of life and trying to figure out which of those claims are true *strictly speaking*, for purposes of philosophical theory. Much of what the ontologist tries to do is to distinguish the genuine logical consequences of things that we believe from merely apparent logical consequences. For example, if we say that there is more than one way to skin a cat, do we *really* mean to say that there are these things, *ways to skin a cat*, such that the number of those things is greater than one? Is that a true logical consequence of what we are saying when we utter that cliché?

In carrying out this enterprise, a big part of what we are trying to do is to figure out which of the words we use correspond to things in the world and which do not. Consequently, some of the technical vocabulary in this chapter and the next has been developed with the goal of clearly distinguishing *linguistic items*—bits of language—from *things in the world* that correspond to those bits of language. Here I want to comment just briefly on some of the more important items in this vocabulary.

Let us begin with the distinction between terms and referents. A term is a word or phrase that refers, or in some other way points to, something in the world. A referent is a thing, a bit of non-linguistic reality, to which the term refers. The name "Socrates" is a term; the *man*, Socrates, is the referent of that term. The phrase "the author of *Metaphysics: The Basics*" is a term; I am the referent of that term. "Spiderman" is a term; the referent of that term is … well, what? It is an interesting philosophical question whether that term has any referent and, if so, what that referent might be. Some terms are not names or descriptions, but are instead predicates, like "is wise". These too can be thought of as having referents. For example, "is wise" might be thought of as referring to the attribute *wisdom*. But this is controversial, partly because of controversies about the nature and existence of attributes, and partly because it is not clear that we would want to say that *reference*, rather than some other relation, is what connects predicates with their corresponding attributes. I will say more about predicates and properties in the next chapter.

Sentences are also linguistic items. Sentences aren't normally thought of as having referents, but many philosophers do believe that typical declarative sentences, and maybe others as well, *express* things in the way that predicates or names do. The things that they express are called propositions. I discuss propositions in the next chapter. (Confusingly, some philosophers in the past have used the word "proposition" to mean something like "sentence". I shall not be using the term in this way. Also confusingly, some philosophers identify propositions with *meanings* whereas others do not. These are controversies that I will mostly ignore.)

Finally, *logic*—or, better, *a* logic—is a system of rules that tells us which sentences follow from, or are consequences of, which sentences. So logic, strictly speaking, governs linguistic items rather than bits of reality; and its rules generally govern "forms of sentences"—or "sentential forms" (or "forms", for short)—rather than their meanings. For example, "Sally is smart and Sprite is a blue unicorn" is a sentence with the form *p and q* where "*p*" and "*q*" are variables that stand for sentences (like "Sally is smart" and "Sprite is a blue unicorn"); and the most familiar and widely used logic—so-called classical logic—tells us that, from any sentence of the form *p and q*, both the sentence *p* and, likewise, the sentence *q* logically follow (i.e., they are logical consequences of *p and q*). So, according to classical logic, "Sprite is a blue unicorn" is a consequence of "Sally is smart and Sprite is a blue unicorn"—never mind the fact that both of these sentences are false.

In the form *p and q*, the variables "*p*" and "*q*", as I have just said, stand for sentences. But the most familiar kind of logic, *standard predicate logic*, also includes variables that stand for predicates and names. So our sentence "Sally is smart and Sprite is a blue unicorn" might also be represented as having the form *Fa* (read: "*a is F*") and *Gb* (read: "*b* is *G*"), where "*F*" and "*G*" stand for predicates like "is smart" and "is a blue unicorn", and "*a*" and "*b*" stand for names, like "Sally" and "Sprite". It will be useful to bear in mind that, according to standard predicate logic, from a sentence of the form "*a* is F" (where "*a*" is a name, like "Sprite", and "F" is a predicate like "is a blue unicorn") a so-called *existential sentence* logically follows, a sentence of the form "*there is something, x,* such that *x* is F" or "*there exists an x* such that *x* is F". Furthermore, the italicized words in each of the latter two

sentences are standardly understood as equivalent acceptable readings of a logical symbol—"∃"—known as the *existential quantifier*; and in standard predicate logic, if you want to express the claim that something is a blue unicorn, that blue unicorns exist, that there are blue unicorns, or anything else relevantly like that, you would simply write "∃*x*(*x* is a blue unicorn)", or "∃*x*B*x*", where "B*x*" abbreviates "*x* is a blue unicorn. In other words, according to standard predicate logic, the following sentences are logically equivalent (i.e., their respective logical entailments are exactly the same): "∃*x*(*x* is a blue unicorn)", "∃*x*B*x*", "something is a blue unicorn", "there are blue unicorns", and "blue unicorns exist". These facts will play an important role in this chapter as we discuss reasons for believing in nonexistent objects.

NONEXISTENT OBJECTS

We often seem to talk about objects that don't exist. There seem, in fact, to be a wide variety of truths about such things. Santa Claus dresses in red; the Easter Bunny does not. Unicorns look like horses, but have a horn. Tyrion Lannister and Katniss Everdeen are more complicated characters than Elrond Half-Elven or Imperator Furiosa. Eleven, from *Stranger Things*, has telekinetic powers very similar to those of force users like Rey and Kylo Ren. And so on. Each of these claims seems *true*. But how can they be true if the names embedded in them do not refer? The claim that Eleven and Rey have very similar powers, for example, seems to be *about* both Eleven and Rey. But how could it be so if Eleven and Rey do not exist (and are not even characters in the same fiction)?

Let us begin our discussion by assuming that fictional characters, like Eleven and Rey, can be treated philosophically like other kinds of nonexistent objects—round squares, unicorns, and the like. The question, then, is what we should say about such things. Shall we say that there *are* nonexistent objects, that we *believe in* nonexistent objects? Do you believe in fairies, leprechauns, unicorns? Should we all agree that there are such things, but that they just don't exist? Or should we insist instead—as we almost certainly would when confronted with someone in the business of ordinary life who believes in such things—that there is *no good reason* to believe in fairies, leprechauns, and so on?

From the point of view of metaphysics, the main motivation for believing that there are things that do not exist—for believing that there are, among other things, fairies, leprechauns, unicorns, and the like—is, as I have just noted, that we seem to talk about and to refer to such things. We seem to talk about fairies and leprechauns, for example, in the course of talking about mythology or works of fantasy. Furthermore, there seem to be generally agreed upon *facts* about fairies and leprechauns. For example, leprechauns typically wear hats and have access to pots of gold. Fairies are generally small and at least some can fly, and they emphatically do *not* have the size, shape, strength, or demeanor of cave trolls. But it is hard to take any of this seriously unless we think that terms like "fairy" and "leprechaun" refer to something. So one way to settle the question whether we should believe in nonexistent things is to determine whether we ought to agree that much of what we *seem* to say and believe about fairies and leprechauns is indeed *about* fairies and leprechauns. If there is good reason to think that, despite initial appearances, we never really say or believe anything that is genuinely about fairies, leprechauns, or other things that do not exist—if the things we seem to say and believe about such things are instead about other things, like the stories we've created, or the contents of our own imagination—then the motivation for believing in nonexistent objects is lost.

We can begin to make some progress on this issue by noting a distinction between two kinds of claims: specific nonexistence claims on the one hand, and a general claim to the effect that there are nonexistent things. 2.1 below is a claim of the first sort whereas 2.2 is a claim of the second sort:

2.1   Pegasus is nonexistent.
2.2   There is something, *x*, such that *x* is nonexistent.

It is easy to see why the truth of 2.1 would give rise to the appearance that we sometimes refer to things that do not exist. Many sentences, particularly those that occur in discussions of Greek mythology, but also sentences like 2.1, seem to refer to Pegasus; and yet Pegasus does not exist. The problem, though, is that there is good reason to doubt that 2.2 follows from 2.1. The reason, as we shall see in a moment, is that there is good reason to think that what one *really* means when

one asserts a sentence like 2.1 is something other than what 2.1 seems at face value to say, just as what one really means when one asks, "What's the matter? Cat got your tongue?" isn't a question about your tongue, or about what any nearby cats might happen to have in their possession. In other words, there is good reason to think that a *paraphrase* of 2.1 does a better job of expressing what 2.1 is really getting at. But if 2.2 doesn't follow from 2.1, then there is good reason to doubt that 2.2 follows from *any* sentence that superficially seems to refer to a nonexistent thing; and if that is right, then it is hard to see what reason we could possibly have for believing that there are nonexistent things. In short, the pressure to agree that there are nonexistent objects comes from the apparent fact that 2.2 follows from sentences like 2.1; but if we can explain away that appearance, then there is no reason to believe in nonexistent things.

So here is our situation: 2.1 looks like a sentence of the form "*a* is nonexistent". As we have already seen, however, standard predicate logic allows us to infer from this the existential sentence "$\exists x(x$ is nonexistent)". But the sentence "$\exists x(x$ is nonexistent)" *cannot possibly be true* on the standard interpretation of the quantifier. This is because, according to the standard interpretation of the quantifier, "$\exists x(x$ is nonexistent)" is logically equivalent to "There exists something that is nonexistent", which is equivalent in meaning to the obviously impossible claim that *something exists that does not exist*. So 2.2 cannot possibly be true. We therefore have three options:

(i)   We can make a distinction between *being* and *existence*, maintaining that there *are* things that don't exist while denying that there *exist* things that don't exist. We can then revise our interpretation of the quantifier accordingly, saying that "$\exists x(x$ is nonexistent)" is logically equivalent to "There *is* something x that is nonexistent", while denying that it is logically equivalent to "There exists something that is nonexistent". Doing this would open up the possibility that 2.2 expresses a truth.

(ii)   We can simply grant that Pegasus exists.

(iii)   We can deny that 2.1 is to be taken at face value and offer a paraphrase. We might say that, although it *seems* to be a claim of the form "*a* is nonexistent", what we *really* mean

to say when we affirm sentences like 2.1—or, perhaps, what we *ought* to mean, or *would* mean to say, if we were thinking clearly about what we are saying and about the logical consequences of what we are saying—is something with a different logical form, something that doesn't imply anything so bizarre as 2.2.

Option (i), as I have just noted, simply grants that there are some things that don't exist. A few philosophers (most notably, Alexius Meinong, Terence Parsons, and Richard Routley) have wanted to take this option; but, even aside from technical problems that arise in trying to develop the view rigorously, it is hard to make sense of the view. Being and existence seem identical; to say that something *is* seems to be exactly the same as saying that it *exists*. If we deny this, then we are under a burden to explain the difference between being and existing, and it is hard to see what that difference could possibly be. Accordingly, options (ii) and (iii) seem generally to be more attractive.

Option (ii) is pretty implausible in the case of Pegasus. But there are interesting reasons for thinking it might be a viable alternative in the case of things that once existed but don't any longer, or in the case of things that could exist but don't in fact exist. Pegasus, it seems, not only doesn't exist, but can't possibly exist. Even if there were winged horses, and even if there was one that had very much the same "life story" as the Pegasus of mythology, no such horse would be the Pegasus of Greek mythology for the simple reason that the myth is not in fact about any horse, winged or otherwise. But consider a different horse, one that did exist but seems not to exist any longer: Bucephalus, the mighty steed that belonged to Alexander the Great. Following a line of reasoning due to Timothy Williamson, we might argue as follows:

2.3   Bucephalus does not exist. (Assume for *reductio ad absurdum*)
2.4   Necessarily, if Bucephalus does not exist, then the proposition that Bucephalus does not exist is true.
2.5   Necessarily, if the proposition that Bucephalus does not exist is true, then the proposition that Bucephalus does not exist exists.

2.6   Necessarily, if the proposition that Bucephalus does not exist exists, then Bucephalus exists.

2.7   Therefore: Necessarily, Bucephalus exists.
         ★★★Contradiction

This argument form is a *reductio ad absurdum* argument: we start with an assumption and reduce it to absurdity by adding premises that, together with our assumption, generate a contradiction. Once we have reached our contradiction, we are allowed to infer that our starting assumption is false. So:

2.8   Therefore: Necessarily, it is false that Bucephalus does not exist.

2.9   Therefore: Necessarily, Bucephalus exists.

If this is right, then not only does Bucephalus *in fact* still exist, but Bucephalus *cannot possibly fail to exist*. Bucephalus is a necessary being. And, of course, the argument generalizes: You and I are necessary beings too, as is everything else. What shall we say about this argument? The only premises that are open to challenge are 2.4–2.6, for 2.3 is just an assumption to get the argument going and the rest are straightforward logical consequences of other premises.

   Premise 2.4 seems pretty obviously true. To see why, suppose there had been no concrete objects whatsoever—nothing exists except maybe abstract objects like numbers, sets, and so on. In that empty world, Bucephalus certainly wouldn't exist. So then it would be true that Bucephalus doesn't exist. But in a world like that, there are no language users, and so no sentences. The only things that could have the property *being true* are the things sentences express—namely, propositions. So the proposition that Bucephalus exists would be true. But a proposition cannot be true unless it exists, since only existing things can have properties like *truth* and *falsity*. So premise 2.5 is true as well. But the proposition that Bucephalus does not exist is *about* Bucephalus, and it can be so only if Bucephalus is "available", so to speak, in order to be the subject of propositions. In other words, the proposition can exist only if Bucephalus does.

   If Bucephalus exists, where do we find him? Nowhere; for, as Williamson points out, agreeing that Bucephalus (still) exists does not

commit us to any view about *how* he exists. On Williamson's view, objects that used to exist but seem no longer to exist are just objects that were once concrete but are no longer. Likewise for merely possible objects, as, for example, you and I would have been if we had never come into existence.

Whatever the merits of this view might be, however, it does not provide us with a fully general solution to the problem of apparently nonexistent things. This is because Williamson's argument only supports the necessary existence of things that (intuitively) *could* exist. If "something" can't possibly exist, then "it" can't possibly be the subject of any proposition, and so there is no reason to think that there is or could have been any such thing as the proposition that it doesn't exist. That does, of course, sound rather weird on the face of things. If my cubical spherical baseball doesn't exist, isn't it *true* that my cubical spherical baseball doesn't exist? And if so, doesn't it just follow that there is such a thing as the proposition that my cubical spherical baseball exists? In short, yes, *unless we can find a suitable paraphrase for the sentence* "My cubical spherical baseball does not exist" that doesn't locate any cubical spherical baseball in the subject position. This brings us to option (iii).

Let us return to the sentence "Pegasus does not exist". What sort of paraphrase can we offer for a sentence like "Pegasus is nonexistent"? Well, that depends to some extent on what we think might be involved in *being Pegasus*, as contrasted with simply *being*, or *being something or other*. Set aside Pegasus for a moment and think about something that actually did exist—our family's pet fish, Sapphibob (so-called because our kids could not agree on whether to name the fish "Sapphire" or "Bob", and so reached a compromise). Adapting a suggestion from W. V. Quine's famous paper, "On What There Is" (1961) we can think of *being Sapphibob* as a kind of activity that something does; and we can then introduce a new verb—*sapphibobbing*—for that activity. To say that Sapphibob used to exist, then, would be equivalent to saying that something used to sapphibob. To say that Sapphibob no longer exists amounts to saying that, at present, nothing sapphibobs. Likewise, then, to say that Pegasus does not exist (and has never existed, and maybe could not possibly exist) would be equivalent to saying that *nothing pegasizes (and nothing ever has done so,*

and maybe it is impossible that anything do so). But, just as one can't infer a claim like 2.2 from the claim that nothing runs, or nothing sits, or nothing plays soccer, so too one can't infer a claim like 2.2 from the claim that nothing pegasizes. For, "Nothing pegasizes" does not have the *form* "*a* is F" and without that form, 2.2 cannot be validly inferred from it. Problem solved.

Alternatively, one might think of *being Socrates* in just the way that property realists—people who, as we will discuss more fully in Chapter 3, believe that predicates like "is wise" correspond to abstract properties, like *wisdom*—think of ordinary predications. So, just as *being red* is equivalent to *having the property redness*, so too *being Sapphibob* might be equivalent to *having the property sapphibobness*. In that case, to say that Sapphibob exists would be equivalent to saying that something exemplifies sapphibobness. Likewise, then, saying that Pegasus does not exist would be equivalent to saying that nothing exemplifies pegasity.

Adopting this strategy amounts to avoiding commitment to nonexistent objects by finding suitable paraphrases of sentences like 2.1 that do not have a logical form that implies an existential sentence like 2.2. It is no part of the strategy to assume that people who assert things like "Santa Claus does not exist" in the ordinary business of daily life really have in mind the paraphrases (in this case, maybe a sentence like "Nothing santa-clauses" or "Nothing has the property *santa-clausness*"). How could they, absent a lot of philosophical reflection or a crash course in the metaphysics of nonexistent things? But the idea is that the paraphrase is something that, at any rate, does justice to such claims and helps us to make sense of how such claims might be true even if terms like "Santa Claus" don't refer to anything.

That said, avoiding commitment to nonexistent objects in this way is not a strategy that comes for free. The price of avoiding commitment to nonexistent objects in this way seems to be commitment to bizarre activities (like pegasizing) or bizarre properties (like pegasity). One might initially doubt that this is much of a cost; after all, as David Lewis once said, "oddity is not falsity". But the bizarreness of the properties and activities in question isn't just their unfamiliarity. Rather, it is that they lack what seems to be one of the defining features of properties and activities—namely, multiple

exemplifiability. Part of what it is to be a property is to be the sort of thing that different objects can, in principle, exemplify, or instantiate; and part of what it is to be an activity is to be the sort of thing that different things can, in principle, do or engage in. So it is not clear what it would be to be a property or activity that can, of necessity, be instantiated or engaged in by just one particular thing. Moreover, it looks as if any specification of the content of a property or activity like this will undermine its suitability to solve the problems it is introduced to solve. For no matter how the content of a property or activity like pegasity or pegasizing is ultimately to be specified, surely the specification would have to make some reference to Pegasus. In other words, pegasity and pegasizing are no less "about Pegasus" than the proposition that Pegasus does not exist; so we should be no less puzzled by the claim that pegasity exists but is unexemplified than we are by the truth of the proposition that Pegasus does not exist. But if that is right, then introducing the property of pegasity cannot solve our problem, and likewise for the activity of pegasizing.

That said, however, many philosophers are nonetheless willing to embrace a solution to the problem of nonexistent things that posits properties or activities like pegasity or pegasizing. Although multiple exemplifiability is widely taken to be a defining feature of properties and activities, it is not hard to imagine that there are special cases where that feature is absent. It is easy enough to understand properties that are not *in fact* multiply exemplified. For example, the property of being the first US president is not (and cannot be) exemplified by more than one thing. To be sure, it *could have been* exemplified by something else; but is it really so hard to imagine another kind of property that not only isn't exemplified by more than one thing but furthermore *couldn't* be exemplified by anything other than one particular thing? Perhaps not. Likewise, from the fact that *we* cannot specify the content of a property like pegasity without reference to Pegasus, it does not obviously follow that the property itself has to be about Pegasus in any deep sense. It is, admittedly, puzzling how its exemplification could guarantee Pegasus's existence if it weren't somehow about Pegasus. But, importantly, neither is it clear how to prove that such a property would have to be about Pegasus.

CREATURES OF FICTION

Let us focus now on creatures of fiction. One might think that in dealing with the problem of nonexistent objects we will automatically settle all questions about creatures of fiction. After all, aren't creatures of fiction nothing more nor less than specific, interesting examples of nonexistent objects? If we reject the latter, don't we automatically reject the former? In a word, no. Although some philosophers treat creatures of fiction as just another kind of nonexistent thing, others insist that there are no nonexistent things but nevertheless admit the existence of creatures of fiction. (In so doing, of course, they deny that creatures of fiction are nonexistent things.) The reason, in short, is that it is harder to paraphrase claims that seem to be about creatures of fiction than it is to paraphrase claims that seem to be about nonexistent objects that are not parts of any particular fiction.

Consider again the case of robustly drawn creatures of fiction—Katniss Everdeen and Imperator Furiosa, for example. Although we can all agree that Furiosa and Katniss are not real people, it is nevertheless easy to cite a great many facts about them, both individually and in comparison with one another. Both are science fiction heroines; both are strong female characters who are able to physically dominate many of the men with whom they come into conflict; both are troubled in various ways; only one (Furiosa) is a character in the Mad Max universe; only one (Katniss) is a minor; etc. We can compare their skills and their armaments. Furiosa has just one arm, and is an excellent shot with a rifle; Katniss has both arms and has superior skill as an archer. We can sensibly ask which would be a better companion or partner in various scenarios; we can speculate about who might win in a fight. It is hard to see how all of this could possibly make sense unless Katniss and Furiosa *exist*. It is hard to see how we could paraphrase these claims in such a way as to say everything we mean while doing away with all of the troubling implications (like *something is identical to Katniss Everdeen* or *something is a creature of fiction*). Similar remarks might be made about Eleven and Rey. How could one paraphrase claims about the ways in which they are similar without doing so in a way that presupposes their existence? So, at least initially, we should be open

to the idea that fictional characters are to be treated differently from nonexistent objects.

Let us sharpen the problem by looking at two particular examples, one which compares two fictional characters by name and another which talks very generally about creatures of fiction without mentioning any by name:

2.10   Katniss Everdeen is younger than Imperator Furiosa, not nearly as well-armed, and exists in an entirely different universe.

2.11   The protagonists of Tolkien novels are much more like mythical heroes than like cyberpunk anti-heroes.

Each of these sentences, like 2.2, *apparently* implies the existence of things (Katniss, Furiosa, protagonists of Tolkien novels, mythical heroes, and cyberpunk anti-heroes) that we commonsensically take not to exist. But how could we possibly paraphrase sentences like 2.10 and 2.11 so as to avoid commitment to such things? It does not seem that we can use the devices of inventing verbs (*katnissizing*, *furiosizing*, etc.) or names for properties (*katnissity*, *furiosity*, etc.) to help us. For it seems quite clear that, whatever we might think about claims like "Katniss does not exist", 2.10 and 2.11 at any rate are *not* about properties like katnissity, furiosity, or whatever properties might stand in for Tolkien protagonists, mythical heroes, and the like. Nor are they about activities like katnissizing and so on. The reason, importantly, is that such properties (if there are any) aren't exemplified, and such activities (if there are any) are not engaged in by anything; so it would be false to say, for example, that something with the property of katnissity is poorly armed, or that something poorly armed is katnissizing.

I will leave it as a challenge to the reader to find suitable paraphrases of 2.10 and 2.11 that do not somehow imply claims like the following existential sentences:

2.12 $\exists x(x = \text{Katniss Everdeen})$

2.13 $\exists x(x \text{ is a Tolkien character})$

If, as is extremely likely, we all fail to find suitable paraphrases, it seems that we face just the same two alternative options we faced

in the case of Pegasus: (i) Admit that Katniss Everdeen, Imperator Furiosa, Tolkien characters, and so on all exist. (ii) Interpret the quantifier as expressing *being* rather than *existence*, and say that Katniss, Furiosa, Tolkien characters, and the like are all nonexistent objects.

If we take the first option, we face the burden of explaining how the claim that Tolkien characters exist is consistent with the commonsense truism that Aragorn, Elrond, elves and hobbits generally, and so on *do not exist*. There are interesting ways of discharging that burden, but to go into them would take us much too far afield. Instead, I want to close by highlighting one other concern for the view that creatures of fiction exist. The concern is that there is no principled way of drawing boundaries between *creatures of fiction* (which, now, we are saying *do* exist) and anything else that we might take not to exist. The reason, in short, is that there are no clear standards for what counts as "creating a fiction".

Earlier I mentioned Quine's well-known paper, "On What There Is". This paper includes an example involving a "round square cupola on top of the Berkeley tower". Intuitively, such a thing does not exist. But if we had a novel which featured a round square cupola on top of the Berkeley tower, such a thing would be a creature of fiction and, on the present view, would exist. But why should writing a *novel* that makes mention of a round square cupola suffice for the creation (or discovery) of such an item when writing a *philosophy article* that makes mention of the same sort of thing does not suffice? Similarly, why should writing a movie script that makes mention of fictional microbes like midi-chlorians suffice for the creation of such a thing whereas writing a (false) chemical theory about, say, phlogiston wouldn't? Why should introducing a character like Katniss Everdeen in the context of a novel suffice for the creation of something whereas introducing a character like Pegasus in a religious myth wouldn't? I can see no sensible answers to these questions. Maybe, then, the case of fictional characters and nonexistent objects are not to be handled so differently after all.

So we end this chapter having discussed two kinds of things that many of us would prefer not to believe in—nonexistent things and creatures of fiction—that at least some philosophers have nonetheless found good reasons for believing in. My goal has not been to convince anyone that they *should* believe in these things. Rather, my

goal has simply been to help us get a clearer picture of *why* philosophers are inclined to believe in and argue about such strange entities. The answer, in short, is that ridding ourselves of them often comes at the price of giving up very commonsensical claims, and believing them sometimes for good reason seems like the best of a bad lot of alternatives. As we shall see in the next chapter, much the same is true for another kind of thing—abstracta—that many philosophers would prefer not to believe in. Our focus in that chapter will be on three kinds of abstract object: propositions, states of affairs, and universals.

FURTHER READING

In the discussion of nonexistent objects, I mentioned work by Alexius Meinong, Terence Parsons, Richard Routley, and Timothy Williamson. Interested readers will want to look at Alexius Meinong, "The Theory of Objects", in Roderick Chisholm (ed.), *Realism and the Background of Phenomenology* (Glencoe, IL: Free Press, 1960), Terence Parsons, "Are There Nonexistent Objects?", *American Philosophical Quarterly* 19, no. 4 (1982): 365–371, Richard Routley, "On What There Is Not", reprinted in Michael Rea (ed.), *Arguing about Metaphysics* (New York: Routledge, 2009)—I will henceforth refer to this book simply as *Arguing about Metaphysics*—and Timothy Williamson, "Necessary Existents", in Anthony O'Hear (ed.), *Logic, Thought, and Language* (Cambridge: Cambridge University Press, 2002). On creatures of fiction, see David Lewis, "Truth in Fiction", reprinted in *Arguing about Metaphysics*, as well as Peter Van Inwagen, "Creatures of Fiction", *American Philosophical Quarterly* 14, no. 4 (1977): 299–308, Stuart Brock, "The Creationist Fiction: The Case Against Creationism about Fictional Characters", *Philosophical Review* 119, no. 3 (2010): 337–364, and Tatjana von Solodkoff, "Explaining Fictional Characters", *Ergo: An Open Access Journal of Philosophy* 6 (2019). For discussion and display of the methodological issues that arose during that discussion, see Quine's "On What There Is", and David and Stephanie Lewis's engaging paper, "Holes", both reprinted in *Arguing about Metaphysics*. See also Peter van Inwagen's "Being, Existence, and Ontological Commitment", in the *Metametaphysics* volume cited at the end of Chapter 1.

ABSTRACT OBJECTS

In the last chapter, we discussed the metaphysics of nonexistent objects and creatures of fiction. In this chapter, we turn our focus to three important kinds of abstract object—propositions, states of affairs, and properties—that many philosophers have wanted to avoid admitting into the catalog of things that exist but that seem like they have to exist if much of our commonsense discourse is to make any sense. I will begin, as before, with some terms and distinctions.

At the beginning of the last chapter, I distinguished terms from referents. I now want to distinguish types from tokens. I have two quarters in my desk drawer. Each is a token of a common type. I said previously that "Spiderman" is a term; but notice that the actual sequence of inked letters between the quotations marks is just a token of a common type. One token of that term occurs in the previous sentence; another token occurs here: "Spiderman". If you read the term out loud, you will produce another token of that term, albeit not a written one. It is important to keep in mind that words for linguistic items—words like "term", "word", "sentence", and so on—can refer to *either* tokens or types. The types, presumably, are abstract objects that have instances; the instances are just the tokens.

In the last chapter I also distinguished names from predicates, and I noted that names have objects as their referents. Whether predicates

also have referents is a matter of controversy. What is relatively uncontroversial is that predicates have extensions, where the extension of a predicate is the **class** of things that satisfy the predicate. Some predicates (like "is a true contradiction") cannot be satisfied; these have the null class (i.e., the class that has no members) as their extension. A class is a collection of objects; and no two classes have exactly the same objects as their members. That is, if a class A has (for example) objects o_1, o_2, and o_3 as its members, and if a class B has those same objects and no others as members, then A is identical to B. This principle is known as the *extensionality axiom*. A set is a class that conforms to further axioms that do not necessarily govern classes. For the most part, the differences between classes and sets do not matter here. But since *class* is the broader category, and because the differences between classes and sets do start to matter as one digs deeper into the details of some of the theories we shall be discussing, I will mostly use the term "class" rather than "set" throughout this chapter.

In addition to having extensions, however, predicates are also sometimes (but not universally) regarded as having properties and relations as their referents. So, for example, "is wise" is sometimes regarded as referring to *wisdom*; and "is paired with" is commonly regarded as referring to the relation of pairing. Properties and relations are sometimes treated as the same kind of thing, where a relation is a property that requires more than one thing in order for it to be exemplified. Wisdom, for example, can be exemplified by just one thing; the pairing relation applies only to pairs of things. The things related by a relation are called the **relata** of the relation.

PROPOSITIONS

One cannot get very far in contemporary metaphysics without encountering talk about propositions, states of affairs, and properties. For many philosophers, these sorts of things are among the most basic building blocks of reality. So having some understanding of them, and of the motivations for believing in them, is crucial for understanding a wide variety of other metaphysical theories. In this section, we focus on propositions.

Friedrich Nietzsche famously said that God is dead. He said it, but he didn't use those words. Nietzsche wrote in German. So his actual

words were "Gott ist todt". Nevertheless, we legitimately attribute to him the saying that God is dead because "Gott ist todt" and "God is dead" are just different sentences that say the same thing. Moreover, not only can we *say* in different languages that God is dead, but speakers of different languages—even those wholly unacquainted with the German language—can believe or disbelieve it, fear or hope for it, regard it as profound or stupid, and so on. Not so, however, with the sentences that express Nietzsche's claim. The sentence "Gott ist todt" cannot itself be said in English. Nor can it be believed or disbelieved, loved or hated, etc. by someone who has no acquaintance with the German language. What this means, then, is that the German sentence, "Gott ist todt" is one thing, and *what the sentence expresses* is something completely different.

We have now identified a role: Being something that is expressed by sentences in different languages, that can be asserted or denied, believed or disbelieved, feared or hoped for, desired or dreaded, and so on. The entities that supposedly fill this role are what philosophers call *propositions*. Propositions have also traditionally been thought of as the fundamental bearers of truth and falsity. What this means is that they are the sorts of things that can be true or false, and everything else that can be true or false derives its truth or falsity from the proposition it expresses. So this is another role that propositions are supposed to play. Combining the two roles, we might say that propositions are the sorts of things that can, in principle, serve as the referents of "that-clauses". (A *that-clause* is any noun phrase beginning with "that".) One can assert *that God is dead*, hope *that God is dead*, say in German *that God is dead*; and it can be true or false *that God is dead*. In each case, "that God is dead" names the proposition that God is dead—or so say the believers in propositions. Of course, owing to the limits of natural languages, some propositions can't *in practice* serve as referents of that-clauses; they might be too complicated, or the relevant words might not exist. But a proposition like this will still be the *kind* of thing that could *in principle* serve as the referent of a that-clause.

Asking whether there are propositions, then, amounts to asking whether anything occupies the roles we have just identified. Asking what propositions are amounts to asking about the nature of whatever it is that occupies those roles.

One might wonder what it means to say that a proposition is an object of attitudes like fear, hope, and so on. After all, nobody is afraid of a proposition; likewise, nobody hopes for one. Here it is helpful to distinguish between two senses of "object of fear". If you are afraid of a rattlesnake in your garden, the snake is an object of your fear in one sense: You fear *it*, or something that it will do to you. This is not the sense in which propositions are objects of fear. Propositions are objects of fear in the sense that they are the *contents* of our fears. If you fear *that the snake will bite you,* the proposition that the snake will bite you is the content of your fear—it is just what you would express if you were to describe in detail exactly what it is that you fear. So likewise they are the contents of our hopes, beliefs, and other such attitudes.

The idea that propositions are objects of attitudes like belief, desire, and so on makes good sense out of a variety of things that we are inclined to say. For example, suppose Alice has been studying late all week with Brian, whom Alice's roommate (Christy) wants to date. Alice hopes, and Christy fears, that Brian would like to date Alice. As it happens, Brian's friends all *believe* that he would like to date Alice. But their beliefs are false; Brian would really prefer to date Christy. It is only natural in this case to say that *what Alice hopes* is at the same time *what Christy fears* and *what Brian's friends all (falsely) believe*. It is hard to make sense of this unless we assume that there is a common object—the proposition that Brian would like to date Alice—toward which Alice's hope and Christy's fear are directed. Likewise, it is hard to make sense of the idea that this common object is *believed* by many people unless we suppose that it is the sort of thing that can be true or false. (Things that cannot have a truth value—things like clouds, or prime numbers—cannot be believed.) Once we admit that it is that sort of thing, it is natural to say that the beliefs of Brian's friends are false *because* the proposition is false; not the other way around. So there seems to be good reason for thinking the roles we have identified are occupied by propositions.

But what are propositions? What is their nature? Given their intimate connection with sentences and beliefs, they would have to be things that could be true or false; they would also have to be things that could stand in logical relations to one another. What you say can be *conjoined* with what I say; it can either be *consistent with* or *contradict*

something I believe; it might *entail* something I believe; and so on. All of this suggests (but does not guarantee) that propositions have something like a linguistic structure—logical form, parts or **constituents** corresponding to subject and predicate, and so on.

There are, in short, three ways in which one might think of propositions: as mental items (beliefs, maybe), as linguistic items, or as abstract objects of some sort. One serious problem with the first two options is that there might never have been any creatures with the capacity for thought or language. But surely in that case it would have been *true* that there are no thinking creatures, that $2 + 2 = 4$, that contradictions can't be true, and so on—and this despite the fact that there would be no bits of language. So if propositions are the fundamental bearers of truth and falsity, they cannot be linguistic items, nor can they be creaturely mental states. One could say that propositions are divine thoughts; but this view commits one to belief in God, which is a commitment many philosophers would be reluctant to take on simply to accommodate a theory of propositions.

We can now see why those who object to abstract objects in general typically also object to propositions. Since there might never have been any concrete things (other than God, perhaps), propositions cannot be identified with anything concrete (outside of God). But thinking of propositions as wholly unique abstract objects is also problematic. Earlier we said that the sentences "God is dead" and "Gott ist todt" *express the same thing*. But how can we possibly know this if propositions are neither concrete objects nor abstract objects of some very familiar sort (like classes, or sets)? If all we know is that propositions are some perhaps very unfamiliar kind of abstract object, we will have no reason to think that we have any real clue as to what rules govern their behavior.

Partly because of this, it is rare now to treat propositions as being wholly different from other kinds of abstract object like states of affairs, properties, relations, and classes. More often people to try to assimilate some of these things to one another. Among the more common theories of propositions, two rise to the fore. The first identifies them with **states of affairs**. The second puts them in the same category as *properties*—either identifying them with properties of a certain kind or else treating properties and propositions as different species of a common genus.

STATES OF AFFAIRS

A state of affairs is a circumstance, a situation, or a way things are. Saying this, however, does not tell us much about what states of affairs are like. It does not tell us much about their intrinsic nature. The two most well-developed conceptions of states of affairs are those of Roderick Chisholm and David Armstrong.

According to Chisholm, states of affairs are necessarily existing abstract entities. They are divided, broadly speaking, into two classes: those that occur and those that do not. My being a philosopher is among those that occur. My running a marathon in less than ninety minutes is, sadly, among those that do not. The latter state of affairs *exists*. I can think about it, talk about it, try (in vain) to bring it about, and so on. It just isn't among the things that have happened or will ever happen in our world's history.

The distinction between *occurring* and *not occurring* mirrors the distinction between *being true* and *being false*. My being a philosopher is among the states of affairs that occur; correspondingly, the proposition that I am a philosopher is true. My running a marathon in less than ninety minutes is among those that do not occur; correspondingly, the proposition that I have, or someday will, run a marathon in less than ninety minutes is false. More generally, it seems that, for every state of affairs that occurs at some time or other, there is a corresponding true proposition and, likewise, for every state of affairs that never occurs, there is a corresponding false proposition. To this extent, then, states of affairs resemble propositions.

There is another point of resemblance. Consider again attitudes like hope and fear. Earlier we said that propositions function as the objects of our hopes and fears, but it makes equally good sense to suppose that states of affairs do. What Alice hopes for and Christy fears, we might say, is nothing more or less than *Brian's preferring to date Alice*. We didn't consider this proposal earlier, because we had not yet introduced anything that might serve as the referent of the phrase "Brian's preferring to date Alice". But now that states of affairs are on the table, we do have a referent. We might say that *Brian's preferring to date Alice* is the name of a state of affairs—one that exists but does not occur. Since it exists, it can be an object of fear and hope. Moreover, it seems that for every proposition that serves as the object of some

attitude like fear, hope, or belief, there will be a corresponding state of affairs that is equally suited to be an object of the same attitude.

Chisholm embraces the idea that states of affairs are the things that we hope for, fear, believe, or deny to be the case. Indeed, on Chisholm's view, the defining characteristic of states of affairs is that they are capable of *being accepted*. But now it is hard to see any good reason for thinking of propositions as distinct from states of affairs. Both are abstract (assuming, again, that we cannot sensibly identify propositions with mental states or linguistic items); both are objects of propositional attitudes; and it is hard to appreciate any real distinction between "occurring" and "being true" or between "not occurring" and "being false". So why not just say that propositions and states of affairs are identical to one another?

In fact, Chisholm does say that all propositions are states of affairs. In particular, he says that they are *eternally occurring states of affairs*. On his view, many states of affairs (e.g., *Sally's eating dinner*) occur at some times but not others; but propositions, he thinks, are either always true or always false. Thus, every proposition is a state of affairs on his view, but not vice versa. So, whereas there is no proposition corresponding to *Sally's eating dinner*, there is a proposition corresponding to *Sally's eating dinner at time t*, which occurs either always or never. Accordingly, the proposition that Sally eats dinner at *t* is either eternally true or eternally false.

The view that propositions do not change their truth values is controversial, but the tendency among philosophers has been to side with Chisholm. One reason is that formal logic is simplified if we assume that propositions have their truth values eternally. Another reason has to do with general metaphysical problems about change over time, some of which will be explored in Chapter 8.

The fact that Chisholm's theory of states of affairs makes it easy to identify propositions with states of affairs is one of its advantages. Even so, Chisholm's theory will not be in the least bit attractive to people who want either to eliminate abstract objects from their ontology altogether, or to identify all abstract objects with classes. One option for those in this latter category is simply to deny the existence of states of affairs. Another is to identify states of affairs with properties or classes. A further strategy would be to identify states of affairs with concrete objects of some sort. This is the approach taken by David Armstrong.

Like Chisholm, Armstrong thinks of states of affairs as the sorts of things that might be named by phrases like *Aristotle's being a philosopher* or *the cat's napping in the sun* or *Christy's liking Brian*. Unlike Chisholm, he thinks of them as concrete entities that have **substances** and attributes as constituents. Familiar particulars—you, me, horses, trees, and so on—are states of affairs, according to Armstrong; and, indeed, he describes the whole world as a "world of states of affairs". In saying this, he does not just mean that the world is populated by states of affairs, but that states of affairs are the fundamental building blocks of the world.

This latter claim might seem to be at odds with the idea that states of affairs have substances and attributes as constituents. After all, if the world is built up out of states of affairs, and substances and attributes are the constituents of states of affairs, then isn't the world *really* built up out of substances and attributes? Not quite. But here we need to say a bit about substances, attributes, and Armstrong's understanding of how exactly they figure into states of affairs.

Very roughly, the distinction between attributes and substances is just the distinction between properties (like *being blue*) and things that *have* properties. Philosophers sometimes use the term "attribute" instead of "property" to signal their adherence to a particular understanding of what properties are like; but I shall use it as synonymous with "property". For Armstrong, attributes are **immanent universals**. To say that they are immanent is to say that they are located in spacetime, in the objects that instantiate them. To say that they are universal is to say that they are instantiated by many objects at once and are therefore wholly located in many places at once (unlike particulars). As for substances, it is a bit tricky to say what these are supposed to be for Armstrong. On the one hand, there is good reason to think that substances are just things like you and me. After all, I have attributes; so do you. So we are clearly things that have properties. On the other hand, it seems rather strange to suppose that *I* am the thing that underlies *all* of my properties. Suppose you strip away properties like *being a philosopher*, *having mass*, *being human*, *having a body*, and so on. What would be left? Certainly not *me*. It seems, instead, that either nothing would be left—in which case there are no substances, but only bundles of properties—or else all that would be left is what philosophers call a "bare particular". In light of this, it is tempting

to think of Armstrong's "substances" as bare particulars—sort of like pin-cushions to which all of the attributes are pinned.

Armstrong is sometimes interpreted as a bare particular theorist. Interestingly, however, his understanding of the way in which substances and attributes are blended in states of affairs calls into question the idea that his substances are just bare particulars. In laying out his view, Armstrong distinguishes between **thin particulars** and **thick particulars**. The thick particular is a state of affairs: specifically, one which has only non-relational properties as constituents. (**Relational properties** are properties that involve relations to other things—e.g., *being ten feet away from a horse* or *having three sides*. Non-relational properties are, as you'd expect, properties that do not involve relations.) The thin particular is an object—so, again, a state of affairs—*considered in abstraction from its properties*. Both count as substances, according to Armstrong. The thin particular is a constituent in a state of affairs; the thick particular has both the thin particular and its properties as constituents. Note, too, that just as a thin particular is a state of affairs considered in abstraction from its properties, so too a universal is (according to Armstrong) "everything that is left in [a] state of affairs after the particular particulars involved in the state of affairs have been abstracted away in thought".

So, for example, consider the state of affairs that consists in Christy's liking Brian. This is a complex state of affairs which includes at least two others—Christy and Brian. The relation *liking* will be a universal, and it is "what is left" when we consider the state of affairs in abstraction from Christy and Brian. Christy and Brian, of course, will be thick particulars. (Christy's liking Brian won't be a particular because it includes a relation among its constituents.) Christy's thin particular will be Christy considered in abstraction from *all* of her properties—so, whatever is left when we remove (in thought) Christy's humanity, her mass, her hair color, her various relational properties (like *liking Brian*), and so on.

In light of all of this, we can now start to see why I say that Armstrong's theory identifies states of affairs, rather than their constituents, as the basic building blocks of reality. The constituents of states of affairs have, on Armstrong's view, no *independent* existence. One never finds thin particulars standing bare without any properties. Likewise, there are no free-standing, uninstantiated universals—no

properties existing apart from any property-bearers. Particulars and universals come only in packages, bound together in states of affairs.

The trouble is that this very feature of Armstrong's theory raises serious questions about the claim that states of affairs have substances as constituents and, indeed, about the claim that they have any constituents at all. If states of affairs are the fundamental entities, then they, not their constituents, would seem to be the things that have the independence requisite for counting as substances. More importantly, it is hard to see how states of affairs as Armstrong characterizes them can even be said to *have* constituents. What is Christy, *considered as the liker of Brian*? Just Christy. What is Christy, *considered as a woman?* Just Christy. What is Christy, *considered as a student*? Still Christy. So what is Christy *considered apart from all of her properties*? Again, Christy. But *Christy* is not a thin particular that exists as a mere constituent of Christy. Christy is, if anything, a thick particular. So it looks as if *there is no thin particular*. Likewise, what is left after we have abstracted away all of the particulars involved in Christy? Well, if the line of reasoning we have just gone through is sound, the only particular involved in Christy is *Christy*; and to consider Christy in abstraction from Christy is to consider nothing at all. So it looks as if there are no universals either. But then it looks as if Armstrong's characterization of states of affairs as substance–attribute complexes can't possibly be correct. For, contrary to what Armstrong says, his view implies that there is nothing contained within a state of affairs answering to the description of either a (thin) substance or a universal.

PROPERTIES

Traditionally, the landscape of views about the nature of properties has been divided into two broad categories: *realist* views, and *nominalist* views. In what follows, I will briefly characterize the distinction between realism and nominalism, and then identify some of the most important versions of each.

In philosophical lingo, **realism** about things of a certain kind is commonly (though not universally) understood as the view that *there are* things of that kind, and that something's being a thing of that kind does not depend on human beliefs, opinions, or concepts.

(I will say more about mind dependence in Chapter 11.) So, for example, given this characterization, most of us are not realists about *leprechauns*, nor are we realists about *obnoxious behaviors*. In the case of leprechauns, we are not realists because we believe that there are none. In the case of obnoxious behaviors, we all know that there are such things (most of us have engaged in them), but we are not realists about them because we see that whether something counts as obnoxious depends on whether people find it offensive. If belching loudly in the middle of a wedding ceremony were generally taken as a sign of respect, it would probably no longer count as obnoxious.

On the traditional conception, properties (if they exist at all) are **universals**, which are abstract objects of a certain kind. And universals (if they exist at all) do not depend for their existence or nature on human beliefs, concepts, or opinions. Accordingly, *realism about properties* is typically identified with the view that there are universals. The alternative view, **nominalism**, is the thesis that universals do not exist. Nominalism is typically identified either with the view that everything is particular (i.e., there are no universals) or with the view that everything is concrete (i.e., there are no abstract objects). The former characterization is the one that is most useful for the present discussion.

Philosophical arguments aside, nominalism is an attractive position. The idea of an abstract object, universal or not, is hard to wrap one's mind around. They are not like us, or like anything else we encounter in nature. They do not *affect* us in any way, so they are even unlike the invisible immaterial things that many of us believe in—God, for example, or souls. Because of all this, many philosophers have thought it would be better not to believe in them if we could avoid doing so. Our overall theory of the world would be simpler, more elegant, and easier to understand if we did not have to suppose that, in addition to the category of familiar concrete entities there is a radically different category of abstract objects.

Much the same can be said about the more specific idea of an abstract universal. Universals are supposed to be objects that somehow lack particularity. But what could it possibly mean to lack particularity? There are no very helpful answers to this question.

Universals are sometimes said to be *present in* many things at once. But what does that mean? They are not in material objects in the way that, say, water is in a cup. They are not ingredients, like tomatoes and basil in pizza sauce. They are not shared among objects like parts can be. So, again, what could we mean by saying that they are *in* the objects that exemplify them? As noted in our discussion of Armstrong, we might observe that universals, in contrast to particulars, *have instances*. Universals are *exemplified*, whereas particulars are not. But this observation does not help much either. It seems that all we really know about having instances or being exemplified is that these are what universals do. So to say that universals but not particulars have instances, or are exemplified, is just to say that universals but not particulars do whatever it is that universals do. We might as well say that universals but not particulars are brillig, and that being brillig is the defining attribute of a universal. We learn nothing from this.

It is, unfortunately, remarkably difficult to avoid believing in abstract universals. The reason, in short, is that it is hard to account for commonsense ways of talking about properties without endorsing realism. We often talk about different kinds of objects sharing features in common. We can pick out our favorite attributes of things. We compare different kinds of attributes in a way that seems to presuppose their reality (e.g., "Bourne's amnesia lends him a human vulnerability that helps to offset his machine-like fighting prowess"). And so on. It is generally thought that if only we can provide a satisfying account of attribute agreement (of specific facts like *x and y both have the property F*, as well as of general facts like *x and y share a feature in common*) we will then also have the resources to account for all of our other commonsense ways of thinking and talking about properties. The challenge to provide an account like this is *the problem of universals.* The particular difficulty that nominalists face is that it is hard to construe all of our talk about features as ultimately being talk about concrete objects and classes (the sorts of things nominalists typically favor). So, say the realists, we ought to give up on nominalism and accept that what we refer to by names like "features", "attributes", and so on are abstract objects belonging to a special, hard-to-define category of their own. That category, of course, is just the category of *universal*.

To appreciate the force of this argument, it will help to take a brief look at some of the main varieties of nominalism. The first thing to note, however, is that *every* version of nominalism either denies that there are properties or identifies them with abstract objects of some sort—usually classes of objects or classes of abstract particularized properties called **tropes** (e.g., *the particular tomato-basil flavor of this spoonful of pizza sauce*). So if you want to avoid believing in abstract objects altogether, you should say that there are no properties. Good luck with that. The most viable versions of nominalism are ones that admit the existence of properties but deny that properties are universals.

There are three versions of nominalism worth highlighting: class nominalism, trope nominalism, and resemblance nominalism. These do not exhaust the territory, but they seem to be the most popular alternatives to realism. We can characterize each of these views by saying briefly how it fills in the blanks in each of the following statements:

(1)   Properties are identical with _____.
(2)   There is a property that x and y have in common if, and only if, ____.
(3)   Something has (e.g.) the property *being human* if, and only if, ____.

In discussing each view, I shall also highlight some of its most salient problems.

Class nominalism maintains, as the label suggests, that properties are classes. Everyone can agree that every predicate is associated with a class of objects—the predicate's extension. (As mentioned earlier, predicates that cannot be satisfied have the null class as their extension.) A natural thought for a nominalist, then, is that the property expressed by a predicate *is* the predicate's extension. A natural further thought is that classes that aren't the extension of any predicate nonetheless count as properties too since, after all, they *could* be associated with predicates if only language were to develop in a certain way. So class nominalists typically believe in a vast array of properties, but reduce them all to classes. According to class nominalism, then, the property *being human* is the class of all human beings, and something

has that property if, and only if, it is a member of the class. Two objects have a property in common on this view if, and only if, there is a class of which both are members.

One of the main attractions of this view is that it identifies properties with abstract objects of a sort that even most nominalists agree that we simply cannot do without. Although many philosophers have wanted to do away with abstract objects entirely, it is widely held that one must at least admit the existence of classes. If one *must* posit abstract objects, so the reasoning goes, then at least we should try to minimize the varieties of abstracta that we posit and stick, if at all possible, to ones (like classes) that are fairly well understood.

There is an obvious problem with class nominalism as I have just presented it, however. As noted earlier, no two classes have exactly the same members. So if a property is nothing other than the class of its instances, then it follows that no two properties have all of the same instances. So, for example, if I were to collect all of the world's ivory billed woodpeckers in my backyard and chase out all of the other birds, then the class of ivory billed woodpeckers would be identical to the class of birds in my backyard. According to class nominalism, then, the *property* of being an ivory billed woodpecker would be identical to the property of being a bird in my backyard. But, of course, that is absurd. Worse, as soon as the ivory billed woodpecker goes extinct, it looks as if class nominalism will imply that the property of being an ivory billed woodpecker is identical to the null class which, in turn, is identical to the property of being a dinosaur (since, of course, the class of all dinosaurs is also the null class). But clearly the property of being an ivory billed woodpecker is different from the property of being a dinosaur.

We might try to avoid the problem by identifying a property with the class whose members are all of its *past, present, and future* instances. Suppose this move is successful. Still, we face another, very similar problem—one posed by properties that have never been and will never be exemplified. Nothing is (or ever will be) a unicorn, a fairy, or a mountain made of gold. Thus, class nominalism will identify all of the following properties with the null class, and so with one another: *being a unicorn, being a fairy*, *being a leprechaun*, and *being a mountain made of gold*. But surely being a unicorn is not the same property as being a mountain made of gold.

There is one more step we can take, however, to minimize the class nominalist's problems. As I will explain more fully in Chapter 4, David Lewis believes that, in addition to the actual world (which comprises all past, present, and future objects), there are infinitely many concrete *possible worlds*. (By "possible" here and throughout the book, except where otherwise indicated I mean **metaphysically possible**.) According to Lewis, a possible world is a universe-sized concrete object. He thinks that for every comprehensive way that a universe *could* be, there is a universe spatiotemporally unrelated to our own where things *are* that way. This implies that there are infinitely many universe-sized concrete objects, all spatiotemporally disconnected from our own universe, and that some of these universes are very similar to ours except in minor details whereas others are vastly and even bizarrely different. Everything that can exist does exist somewhere in the space of possible worlds. Thus, just as the class of human beings has members, so too the class of unicorns, the class of fairies, and the class of mountains made of gold have members. So we can avoid the aforementioned problems by identifying a property not with the class of all of its *actual* (past, present, and future) instances but with the class of all of its *possible* instances. The property *being human*, for example, is just the class that includes every human being from every possible world. The property *being a unicorn* is the class that includes every unicorn from every possible world. And so on.

One advantage to Lewis's view is that, as noted in an earlier section, it identifies *propositions* with classes. On Lewis's view, every proposition is identical to the property of being a world at which that proposition is true. (Well, almost. Certain special kinds of propositions require special treatment. But we shall ignore those complications here.) So every proposition is a property, and every property is a class. For example, the proposition that you are studying metaphysics is just the property *being a world where you are studying metaphysics* which, in turn, is just the class that includes every world in which you are studying metaphysics. This is a nice and tidy view in which apparently disparate kinds of abstract objects are reduced to a single kind—classes.

But tidiness comes at a price. For starters, note that you can accept Lewis's view only if you believe that there are unicorns, golden

mountains, leprechauns, and the like. You do not have to believe that there are such things in *our* universe, of course; but Lewis's view implies that such things do exist, albeit spatiotemporally unrelated to us. Do you believe in such things? If not, look elsewhere for a theory of properties. If so, consider the further consequences of Lewis's view and see if you can accept those as well.

Like other versions of class nominalism, Lewis's view maintains that necessarily co-extensive predicates (i.e., predicates that are necessarily such as to have the same extension) express the same property. For example, since it is a necessary truth that every triangle has three sides and vice versa, the class of all possible triangular things is identical with the class of all possible trilateral things. So Lewis's view implies that the property of being triangular is identical to the property of being trilateral. So far, one might not be too concerned. This consequence is, at any rate, not nearly as bad as saying that *being an ivory billed woodpecker* is the same property as *being a bird in my backyard*. But the problem worsens when we turn our attention to sentences that cannot possibly be true, and predicates that cannot possibly be satisfied. These all will express the null class, which means every unsatisfiable predicate expresses the same thing as every sentence that can't possibly be true. So Lewis's view implies (for example) that the property *being both a man and a sugar cube* is identical to the property *being a gold mountain made entirely of water* and to the proposition expressed by "Bill is a married bachelor".

Finally, since necessary truths are propositions that are true in all possible worlds, Lewis's view implies that, if a sentence expresses a necessary truth, the proposition it expresses is the class of all possible worlds. So "2 + 2 = 4" expresses the same proposition as "it is morally wrong to torture small children for fun", and this proposition, oddly enough, is identical to the property of *being a possible world*. Add all of these costs together and it is easy to see why Lewis's brand of class nominalism—which seems to be the most viable brand—is widely rejected.

There is one further objection against class nominalism that we should discuss before moving on, because it helps to make our transition to resemblance nominalism. According to class nominalism, having a property is *simply* a matter of belonging to a certain class.

There is nothing in the story that indicates *why* one class rather than another is associated with a given predicate. A realist might say that the class of humans is the extension of the predicate "is human" because *all of the members of that class instantiate a common universal—the one we call "humanity"*. But the class nominalist has no explanation to offer. This is puzzling.

Resemblance nominalism does not suffer from this last problem. According to the resemblance nominalist, attribute agreement is explained ultimately by appeal to a primitive (unanalyzable) resemblance relation. If there are properties, then they are classes whose members resemble one another, or certain paradigms. The nature of the property is defined by the paradigms or by the nature of the resemblance among the members of the class. So some resemblance nominalists will say that something has, e.g., the property *being human* if, and only if, it resembles paradigm human beings or, if there are no paradigms, resembles everything else that belongs to the class of all human beings. There is a property that two things have in common if, and only if, those two things are members of a common resemblance class.

Another advantage to resemblance nominalism is that it avoids the problem of uninstantiated and necessarily instantiated properties. Although every predicate will have an extension, there is no pressure to say that all predicates express properties, since there is no guarantee that the members of every predicate extension will genuinely resemble one another. Not only is this advantageous from the point of view of trying to avoid problems associated with uninstantiated and necessarily instantiated properties, but it is also advantageous insofar as it allows the resemblance nominalist to endorse what is known as a *sparse* theory of properties. A **sparse theory of properties** maintains that relatively few predicates correspond to properties. The motivation for such a theory is just the idea that not every predicate corresponds to a genuine mode of resemblance. For example, infinitely many predicates are satisfied by both goats and microwave ovens. Goats and microwaves are such that seven is a number, they are not trees, etc. But it would be odd to say that, by virtue of this fact, a goat and a microwave *resemble* one another. One way of explaining why this is odd is to say that *being such that seven is a number* and *being*

a non-tree are not genuine modes of resemblance—they are not real properties. One who endorses this sort of view will say that a great many predicates (not just the paradoxical ones) fail to express properties; hence the description of such a view as a "sparse" theory of properties. Again, one advantage to resemblance nominalism is that it can accommodate such a view.

But, as one might expect, there are also significant problems. First, note that we still face the problem of *co-extensive* properties. The resemblance class that includes all and only triangles is (of necessity) identical to the resemblance class that includes all and only trilaterals; so resemblance nominalism has no way of distinguishing these two properties. But, intuitively, the property of having three *sides* is distinct from the property of having three *angles*.

Second, we might well ask where the paradigms are. Who are the paradigm humans, for example? What are the paradigm red things? Any answer, it seems, will be hopelessly arbitrary. For this reason, resemblance nominalists now typically refrain from saying that property-having depends on resemblance to paradigms; rather, they say that having a property is simply a matter of belonging to a resemblance class. For example, being human is not a matter of resembling paradigm humans but is, rather, a matter of belonging to a certain class of things (the class of human beings) all of which resemble one another.

But doing away with the paradigms raises a third problem, the so-called problem of *imperfect community*. Suppose, for example, we add some dogs, crocodiles, and a variety of humanoid robots to the class of human beings. We would still have a resemblance class, but we would no longer have a class that corresponds to a property. Why? Because, although all the objects in this new class would still resemble one another, they would not all resemble one another in the same way. What we want, then, is a way of distinguishing different *ways* in which objects resemble one another: a dog and a human being resemble one another in one way, a human being and a humanoid robot resemble one another in a different way, a humanoid robot and a crocodile resemble one another in yet a third way, and so on. We could do this by supposing that there are a great many different resemblance relations (one for each property). For example, we could say that some objects *human-resemble* each other whereas others

mammal-resemble each other, and so on. But having to say something like this seems to greatly diminish the attractiveness of resemblance nominalism, and it is hard to see other viable options.

Finally, we turn to *trope nominalism*. Trope nominalism identifies properties with sums or classes of tropes. Tropes, as I said earlier, are supposed to be abstract particular entities like *the particular whiteness of my shirt*, *Socrates's humanity*, etc. Tropes are typically referred to as "property-instances", presumably because they are tokens of particular property types. (Just as a particular quarter in my desk drawer is a token of the type *quarter*, so too the particular tomato-basil flavor of a spoonful of pizza sauce is a token of the type *tomato-basil flavor*.) Furthermore, trope theorists typically say that familiar objects are somehow built up out of tropes. So something has, e.g., the property *being human* if, and only if, it has a *humanity* trope among its constituents. What it is for two objects to share a property in common—being the same color, for example—is just for each object to have a constituent—a trope—that *perfectly resembles* a corresponding constituent of the other object.

Trope theory avoids the problems associated with class nominalism in much the same way that resemblance nominalism does. Trope theorists face no pressure to admit the existence of uninstantiated properties, since there is no reason to think that every predicate extension will have members that all share a property in common. Moreover, trope theory also avoids the main problems associated with resemblance nominalism. Necessarily co-extensive properties can be distinguished by appeal to differences among tropes. For example, we have no good reason to think that a three-sidedness trope would perfectly resemble any triangularity trope; so there is no obvious obstacle to insisting that they would differ from one another. Likewise, since properties are classes of perfectly resembling tropes, there is no problem specifying paradigms, nor is there any apparent way of generating a problem of imperfect community.

So trope theory has some real advantages. But, like the other versions of nominalism, it too has some serious drawbacks. For example, consider the properties *of tropes*. Tropes obviously have properties: *being abstract* and *being particular* are the most notable. But it would be absurd to say that all tropes have *abstractness* and *particularity* tropes as parts. For one thing, tropes are supposed to be

pure property instances—they don't have parts. Moreover, it seems odd to say (as we would have to) that abstractness and particularity tropes themselves have abstractness and particularity tropes as parts. Indeed, this would imply that every abstractness trope is infinitely complex: it would have an abstractness trope as a part which, in turn, would have another abstractness trope as a part, and so on.

Of course, I have already acknowledged that trope theorists need not say that every predicate corresponds to a property. Why not apply that strategy here and deny that "is abstract" and "is particular" express properties of tropes? The answer is that if one did so then one might as well deny the existence of properties altogether. As we have already noted, some predicates—e.g., "is such that the moon exists and $2 + 2 \neq 22$" or "is a unicorn"—do not seem to pick out genuine modes of resemblance among things. It makes sense to deny that such predicates express properties. But tropes *do* resemble one another by being abstract and particular—indeed, the kind is defined by way of the predicates "is abstract" and "is particular". We can thus sensibly deny that those predicates express properties only if we can sensibly deny that *any* predicates that pick out genuine modes of resemblance express properties. One can try to give up on properties altogether; but, as noted earlier, there seems to be little reason for optimism that such a strategy can work.

There is another problem for trope theory worth mentioning as well. Consider the particular whiteness of my shirt. Call that trope "T1". According to trope theory, my shirt counts as white by virtue of having T1 as a part. But now consider two other objects: the left half of my shirt and the right half. These are distinct objects that resemble one another perfectly with regard to their whiteness. Each, then, must have its own distinct whiteness trope, T2 and T3, both of which will perfectly resemble T1. Now consider four other objects, namely the top and bottom halves of the left and right halves of my shirt. Each of *these* objects is white too, and so each will have a whiteness trope distinct from T1–T3. In the region occupied by my shirt there will be a myriad perfectly resembling and substantially overlapping whiteness tropes. But what grounds, really, do we have for believing in T1 anymore? Trope theory posits T1 to explain *the whiteness of my shirt*; but my perception of whiteness in the region occupied by my shirt, as well as the resemblance between what I see

when I look at the shirt and what I see when I look at other white shirts is perfectly explained by the presence of the myriad "smaller" whiteness tropes that are present in the smallest white portions of my shirt. T1 is explanatorily and perceptually redundant. Moreover, the same goes for T2, T3, and, indeed, for every whiteness trope that overlaps the smallest whiteness tropes present in the region occupied by my shirt. So it seems that we should not believe in T1, T2, and so on. But if there is no such thing as *the particular whiteness of my shirt* then, according to trope theory, my shirt is not white. So the problem comes to this: It seems as if my shirt can be white simply by having a sufficiently large number of small parts that each have their own whiteness tropes; but trope theory insists that *nothing* is white unless it has its own distinct whiteness trope.

Here, as with the other two varieties of nominalism, I do not take myself to have given an exhaustive list of objections, and neither do I take the objections to be decisive. But the objections are serious enough and the responses baroque enough that realism deserves a serious hearing. The main problem with realism seems just to be the outright weirdness and elusiveness of the universal–particular distinction. However, when we compare this weirdness with the weirdness of saying that the null class is both a property and a proposition, or that there are infinitely many distinct resemblance relations, or that there are tropes that somehow entirely lack properties, it is not so clear that realism comes up with the short end of the stick.

As with nominalism, realism comes in several varieties. But here we can be much more brief, for the varieties are easily characterized by reference to three main decision points, and there is no need to identify multiple distinct problems with each version of realism. For, again, the main problem for all versions is essentially the same—all are committed to a universal–particular distinction, and to the view that the category of universal and the category of particular both have instances.

The first decision point concerns the question whether universals are *immanent* or *transcendent*. Plato, Bertrand Russell, and many others think of universals as transcendent in the following sense: They are wholly outside space and time, and they are not in any way "present in" the objects that exemplify them. Often the term "participation" is used to characterize the relationship between particulars and

transcendent universals; often, too, people draw a sharp distinction between "exemplification" and "instantiation" (the former being the relation between particulars and transcendent universals, and the latter being the relation between particulars and immanent universals). But it is not clear that there is any substantive content to terms like "participation" and "exemplification" beyond their contrast with terms that are supposed to indicate the *inherence* of universals in particulars. As you might guess from what has just been said, believers in immanent universals think that universals are somehow present in their instances. Some go so far as to say that they are parts of their instances. Aristotle is among the defenders of immanent universals and, as we have already seen, so is David Armstrong.

The second decision point concerns an issue that also arises in connection with nominalism: Do we endorse a sparse theory of properties or an *abundant* or *plentiful* one? That is, do we endorse a restrictive view about which predicates correspond to properties, or do we say that just about every meaningful predicate expresses a property? Peter van Inwagen, for example, opts for an abundant theory. According to van Inwagen, properties are *assertibles*—things that can be said of other things. Since every meaningful predicate expresses something that can be said of something else—that's just what predicates do, after all—it turns out, on his view, that every such predicate expresses a property. David Armstrong, on the other hand, believes that quite a lot of predicates fail to express universals. His view, rather, is that universals constitute the "genuine" modes of resemblance among objects, and the mere fact that two objects satisfy the same predicate by no means guarantees that the two objects genuinely resemble one another. Returning to an earlier example: You and the number 77 both satisfy the predicate "is such that the moon exists and $2 + 2 \neq 22$". Does it follow from the fact that you both satisfy this predicate that you *resemble* the number 77? Armstrong and many others will say no—you and the number 77 are not similar at all.

Lastly, we need to decide whether we believe in *unexemplified* or *uninstantiated* universals. Obviously this decision is related to our previous one. If, for example, we endorse a view like van Inwagen's, we can hardly avoid saying that there are unexemplified properties. For, after all, among the things that can be said of me are *that I am a*

unicorn, *that I am a round square*, and so on. Of course, these things are falsely said of me (if said at all); but they can be said of me nonetheless. Others, however—Aristotle and Armstrong among them—endorse a *principle of instantiation*, according to which there are no uninstantiated universals. On this view, then, "is a unicorn" does not express a universal, even though it would have done so had there been unicorns. One consequence of this view is that most universals exist contingently, contrary to what many realists are inclined to think. If there had been no elephants, for example—and it seems that there could easily not have been—then there would have been no property of *being an elephant*. One attractive feature of this view is that it allows us to say that universals depend in an important way on the things that have them, rather than the other way around.

FURTHER READING

An excellent starting place for some of the issues in philosophy of language about meaning and reference that were touched on in the section entitled "Preliminaries" is Jeff Speaks's encyclopedia article, "Theories of Meaning", in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, 2011, available online here: http://plato.stanford.edu/archives/ sum2011/entries/meaning/.

The theories about propositions, properties, and states of affairs that were attributed to David Armstrong, Roderick Chisholm, and David Lewis can all be found in the following works by those philosophers: D. M. Armstrong, *A World of States of Affairs* (Cambridge: Cambridge University Press, 1997); Roderick M. Chisholm, *Person and Object: A Metaphysical Study* (La Salle, IL: Open Court, 1976); David K. Lewis, *On the Plurality of Worlds* (Oxford: Blackwell, 1986). David Lewis, whose work is discussed in most of the chapters of this book, was one of the most influential and systematic metaphysicians of 20th century. A good resource for understanding his overall system is Daniel Nolan's, *David Lewis* (Montreal: McGill Queens University Press, 2005).

In the section on propositions, I raised a question about whether it makes sense to treat propositions as objects of attitudes like fear and hope. For more detailed discussion of that issue among others, see Trenton Merricks's book, *Propositions* (Oxford: Oxford University Press, 2015). On the subject of structured propositions, I have learned a great deal from Lorraine Keller, "Whence Structured Propositions?" (PhD Thesis, University of Notre Dame, 2012). Interested readers may also wish to consult *New Thinking about Propositions*, edited by Jeffrey C. King, Scott Soames, and Jeff Speaks (Oxford: Oxford University Press, 2014).

For a very accessible book-length treatment of the issues about properties discussed in this chapter, see D. M. Armstrong, *Universals: An Opinionated Introduction* (Boulder, CO: Westview Press, 1989). Bertrand Russell's "The World of Universals", pp. 91–100 in *The Problems of Philosophy* (Oxford: Clarendon Press, 1912) is a classic statement of realism about universals. Peter van Inwagen's "Theory of Properties", and David Armstrong's "World of States of Affairs" (not to be confused with his book of the same title) also develop realist views about properties, and both are anthologized in *Arguing about Metaphysics*. Henry Fitzgerald's "Nominalist Things", also anthologized in *Arguing about Metaphysics* provides a light-hearted characterization of the distinction between "nominalist-friendly" objects and the sorts of things that only a realist would believe in. H. H. Price, "Universals & Resemblances", pp. 7–32 in *Thinking and Experience* (Cambridge, MA: Harvard University Press, 1953) provides a classic statement of resemblance nominalism. A classic statement of trope theory is D. C. Williams, "The Elements of Being", *Review of Metaphysics* 7, no. 2 (1953): 3–18, 171–192.

POSSIBLE WORLDS

It is commonplace to talk about "the world" as if there is just *one* world—namely, the one four-dimensional universe in which, as it were, we live and move and have our being. But many philosophers and scientists believe that, in fact, there are many things that deserve to be called worlds. Some believe in parallel universes that together comprise a multiverse. Some believe that even within our universe there are, at any given time, many branching "paths" along which the future might unfold, and that these paths are no less concretely real than the events you are presently experiencing. Some believe in special kinds of abstract states of affairs that constitute complete, alternative ways in which our universe (or multiverse) might have turned out. All of these kinds of things and more have been characterized as *other worlds;* and, what's more, metaphysicians commonly identify some of these other worlds as *merely possible* worlds.

There are many different reasons for believing in other worlds of various kinds. Some do so because they think that the existence of such things is probable on the assumption that spacetime is infinite. Some posit other worlds to explain certain features of quantum mechanics. Some say that, if we don't believe in other worlds, then it appears that certain features of our universe can only be explained by

supposing that a cosmic designer exists. These motivations are commonly characterized as *scientific* reasons for believing in other worlds. The issues are interesting, not just for metaphysics and science but also because of their connections with the so-called "fine tuning" argument for God's existence. But I shall not pursue them here. The metaphysics of possible worlds, and philosophical reasons for believing in them, are the focus of this chapter.

Philosophers posit possible worlds primarily as a way of making sense of our talk about possibility and necessity—about how things *could have been* and how things *must be*. Philosophers sometimes distinguish different kinds of possibility and necessity: physical, logical, metaphysical, and so on. I shall say a bit more about these distinctions later. For now, I'll simply note that when I talk unqualifiedly about possible worlds, or about what is simply possible or necessary, it is generally metaphysical possibility and necessity that I have in mind.

As I will explain in more detail below, most philosophers think that possible worlds are abstract states of affairs. However, some philosophers do not believe in abstract possible worlds, but instead think that certain kinds of parallel universes deserve to be called possible worlds. If they are right, then possible worlds are concrete objects. Part of our task, then, in what follows will be to flesh out these two different conceptions of possible worlds more fully and to explore some of the reasons for and against endorsing each one.

MODALITY

The English language has a variety of **modal expressions**: *may, might, must, can, is possible, is necessary,* and so on. Philosophers posit possible worlds in order to explain what we mean by these expressions. But before discussing different concepts of possible worlds and the role possible worlds play in helping us to make sense of our modal expressions, we should first make some very basic distinctions and explain some of the terminology that will be central to our discussion.

Philosophers make distinctions between logical, physical, and metaphysical modality. But there are other forms of modality as well—deontic (pertaining to morality) and epistemic (pertaining to what we believe or know), for example. Roughly speaking, the various

kinds of modality are distinguished from one another by the various kinds of "laws" relative to which they are supposed to hold: what is physically or **nomologically** possible is what is permitted by the laws of nature; what is logically possible is whatever is permitted by the laws of logic; what is epistemically possible is whatever might be true given what you believe (or know) about the world; and that which is *morally permitted* is the analog of "possibility" in the deontic realm. Accordingly, we might go on to say that what is metaphysically possible is whatever is consistent with the "laws of metaphysics" (whatever those might be). Alternatively, we might just say that what is metaphysically possible is just whatever is possible in the broadest, most all-inclusive sense of the term. As I said earlier, whenever I talk unqualifiedly about possibility or necessity, it is metaphysical possibility and necessity that I have in mind.

Not all of these forms of possibility coincide with one another. It is obvious that metaphysical possibility diverges from deontic and epistemic possibility. For example, plenty of things are possible but not morally possible; and plenty of things that might be true for all we know are nonetheless metaphysically impossible. Furthermore, many philosophers think that logical, physical, and metaphysical possibility diverge. If the laws of nature could have been different then what is metaphysically possible is broader than what is physically possible. If some logically consistent proposition could not have been true, then the fact that something is logically possible doesn't imply that it is possible, period. For example, many agree that the proposition *that God has done something wicked* is not logically contradictory. But one might still think that it could not possibly be true. I shall not try to defend the conclusion that these different modalities diverge. I shall continue to assume that they do, but readers should bear in mind that this is not a universally shared assumption.

Pretty much everyone agrees that modality attaches to propositions. Some propositions are necessarily true, some are necessarily false, some are contingently true, some are contingently false, some are possibly true, and some are possibly false. These modal expressions are interdefinable. A proposition is necessarily true if, and only if, its negation is not possibly true. A proposition is contingently true if, and only if, it is true but not necessarily true. A proposition is contingently false if, and only if, it is false but not necessarily false.

A proposition is possibly true if, and only if, its negation is not necessarily true. And so on. Many philosophers have hoped that modal expressions could be defined solely in terms of non-modal expressions. That is to say, many have hoped that the modal could be *reduced* to the non-modal. But so far that hope has not panned out. Thus, people sometimes speak of "the modal circle" to convey the idea that, although modal expressions can all be defined in terms of one another, ultimately the definitions just go in a circle; there is no way to define the modal in terms of the non-modal.

Many philosophers also think that at least some of the properties of objects have modality built into them. We call these **modal properties.** Examples include properties like *being able to survive squashing*, or *being necessarily (or essentially) a cat*. But the idea that objects have modal properties is a lot more controversial than the idea that modality attaches to propositions. When philosophers wish to talk about modality construed as attaching to propositions, they speak of *de dicto* modality, where "*de dicto*" means something like "pertaining to what is said". When they wish to talk about modality construed as built into the properties of objects, they speak of *de re* modality, where "*de re*" means something like "pertaining to the object".

Finally, it will be helpful to introduce some of the logical symbols that enter into discussions of possibility and necessity and which are, accordingly, scattered liberally throughout the literature in contemporary metaphysics. Following standard conventions, I shall use "□" to mean "necessarily" and "◊" to mean "possibly". The box and diamond are called *logical operators*: we attach them to sentences (like *God exists* or *Sally stands up*) to form new sentences (like *Necessarily, God exists* or *Possibly, Sally stands up*). We'll use this symbol for strict implication, or entailment: "⇒". To say that p strictly implies, or entails, q is just to say that it is a necessary truth that if p is true, then q is true.

THE UTILITY OF POSSIBLE WORLDS

So much, then, for preliminaries. Now we can turn to our discussion of possible worlds. Possible worlds were introduced to help explain what modal operators mean. Of course, in one sense we already know what they mean: the box means "necessarily"; the diamond means

"possibly". But in another sense people had no real clue what they meant. To see why, just consider the fact that we can keep on adding operators to the beginnings of our sentences to form new (meaningful!) sentences. Thus, consider a sentence like, "□(God exists)". Suppose you know what this means. Still, we can form a new sentence by adding another operator out front, as follows: "◊□(God exists)". Is it clear what *this* means? If so, how about this: "◊◊□(God exists)"? Or this: "◊◊□◊□(God exists)"? If all you know about the box and the diamond is that they mean "necessarily" and "possibly", you will have a very hard time making sense of these sentences.

Now you might wonder, *who really cares about weird sentences like these?* Even if they are meaningful, their meaning will surely be something highly abstract and so not very useful for our lives (or even for serious philosophy). But the fact is, there are several different formal systems—the different logics—that logicians have developed to try to express the sound principles of reasoning about necessity and possibility; and you will sometimes reach very different conclusions in metaphysics depending upon which of these logics you accept. To take just one example, here is an argument for the existence of God that is valid in one modal logic (the one that most metaphysicians accept, as it happens) but not in many of its rivals:

4.1   ◊(God exists)
4.2   ◊(God exists) ⇒ ◊□(God exists)
4.3   Therefore: ◊□(God exists)
4.4   Therefore: □(God exists)

4.1 looks like an innocent assertion of possibility. Of *course* it is possible that God exists. 4.2 is supposed to be a conceptual truth about God. The concept of God is supposed to be the concept of a *necessary* being—a being who couldn't fail to exist. So if it is possible that God exists, it is possible that God is a necessary being; hence, it is possible that God exists necessarily. So 4.3 follows from 4.1 and 4.2. But then comes another inference—the one from 4.3 to 4.4. Is *that one* valid? It is if the following principle is true:

4.5   ◊□p ⇒ □p

But how might one begin to assess this principle? Again, if all we know is that "◊" means "possibly" and "□" means "necessarily", it is very hard to tell. What one needs in order to facilitate understanding and to make proper evaluation of claims like 4.5 possible is a *semantics* for the modal operators—an account of what those operators mean.

In 1963, Saul Kripke proposed what was to become the standard semantics for the modal operators. According to Kripke's interpretation, sentences with modal operators in them are sentences about other possible worlds. Thus, □*p* is true if, and only if, *p* is true in *every* world that is possible relative to ours, and ◊*p* is true if, and only if, *p* is true in *some* world that is possible relative to ours. Another way of putting this is to say that the box and the diamond are *quantifiers* over possible worlds. ◊*p* is true if, and only if, *there exists a world* that is possible relative to ours in which *p* is true. □*p* is true if, and only if, *for any world* w *that is possible relative to ours, w* is a world in which *p* is true. This way of interpreting the symbols helped philosophers to get an intuitive grip on the differences among various rival systems of modal logic.

To appreciate this a little better, just imagine a space of possible worlds. Let's think of worlds as concrete parallel *universes*—i.e., spacetimes that are not spatiotemporally related to ours. The one we live in is the actual world; the others represent other possibilities for how the universe might have gone. Intuitively, then, every one of those universes counts, from our perspective, as *a way things might have gone*. Furthermore, suppose that for *every* possible way things might be, there's a parallel universe in which things are just that way. Some of the universes have different laws of nature; thus, our universe might have had different laws of nature. None of them have spherical cubes; so there being a spherical cube is not possible—it is not included in any way things might have gone. And so on.

Now, ask yourself this question: From the perspective of one of those other universes, is *ours* a way things might have gone? And is every universe that is possible from the point of view of one of those other universes *also* possible relative to ours? If you say yes, then you are probably thinking of the space of possible worlds as analogous to a classroom full of students, and the "relative possibility" relation as being analogous to the *is in the same classroom as* relation. Everyone is in the same classroom as themselves (which is to say that the relation is **reflexive**). Furthermore, if *x* is in the same classroom as *y*, then *y*

is in the same classroom as x (so, the relation is **symmetric**). Also, if x is in the same classroom as y and y is in the same classroom as z, then x is in the same classroom as z (so, the relation is **transitive**). Once we have arrived at this way of thinking about the relative possibility relation, we can now evaluate 4.5. Consider someone in the same classroom as me. Now suppose that, from their point of view, *everyone in the classroom* is wearing a red shirt. Well, then it has to be the case that, from *my* point of view too, everyone in the classroom (including me) is wearing a red shirt. So likewise: Consider one of the worlds possible relative to ours. Now suppose that, from the point of view of that world, *every* possible world is one in which p is true. It follows, then, that from the point of view of our world too, every possible world (including ours) is one in which p is true. But this is just to say that 4.5 is true.

On the other hand, suppose you think differently about relative possibility. Suppose that when you considered the perspective of one of the worlds possible relative to ours, you thought like this: Maybe from the point of view of *that* world, there are ways things could have gone that simply aren't, from our point of view, ways things could have gone. In that case, you are probably thinking of the relative possibility relation as analogous to the *has an acquaintance in common with* relation (on the assumption that everyone has at least one acquaintance, if for no other reason than that everyone is acquainted with themselves). This relation is both reflexive and symmetric: everyone has an acquaintance in common with themselves, and if x has an acquaintance in common with y, then y has an acquaintance in common with x. But it is not transitive. That is, it is not generally true that if x has an acquaintance in common with y and y has an acquaintance in common with z then x has an acquaintance in common with z. So consider someone who has an acquaintance in common with me. Now suppose that everyone who has an acquaintance in common with *them* has read *Lord of the Rings.* Does it follow that I have read *Lord of the Rings*? Yes. But does it follow that everyone who has an acquaintance in common with *me* has read *Lord of the Rings*? No, since maybe some of the people who have acquaintances in common with me don't have acquaintances in common with them. So likewise, if relative possibility behaves like the *has an acquaintance in common with* relation, 4.5 will be false.

Understanding possibility in terms of possible worlds helps us in evaluating claims like 4.5, in part because it helps us think more clearly about whether we ought to say that the relative possibility relation is reflexive, symmetric, transitive, or some combination of these. Thus, it helps us to determine which of the different modal logics we want to accept. Once we have adopted this framework for understanding the modal operators, however, we have to ask: what are these possible worlds that we are talking about? There are various answers to this question, but I want to single out two for consideration here: (i) concrete objects and (ii) abstract states of affairs.

CONCRETE POSSIBLE WORLDS

David Lewis recommends thinking about possible worlds as something like parallel universes. We can state his understanding of possible worlds a bit more accurately, however, as follows. Let us say that two things are *worldmates* if, and only if, they are spatiotemporally related to one another. According to Lewis, the *actual world* is a universe-like concrete object whose parts are *us and all of our worldmates* and nothing else. (In fact, maybe it just *is* our universe—depending, of course, on how we define "universe".) Moreover, Lewis also thinks that there are other things that are not our worldmates. So now consider one of these things—call it *Fred*. According to Lewis, there is another world, distinct from ours, whose parts are just *Fred and Fred's worldmates*. This too will be a universe-like concrete object; and (like Fred) it will not be spatiotemporally related to us. And so on. For every object that is not spatiotemporally related to us, there is a universe-like object whose parts are it and all of its worldmates. Each of these universe-like objects is a world; and, by definition, no two worlds are spatiotemporally related to one another. According to Lewis, there are infinitely many worlds, enough to represent absolutely every possible way things might have been.

Why believe in concrete possible worlds? We have already seen one reason to believe generally in possible worlds: Doing so gives us the resources to understand modal logic. In his book, *Counterfactuals,* Lewis gives another reason. Surprisingly, it is an appeal to common sense:

> I believe, and so do you, that things could have been different in countless ways. But what does this mean? Ordinary language permits the

paraphrase: there are many ways things could have been besides the way they actually are. I believe that things could have been different in countless ways; I believe permissible paraphrases of what I believe; taking the paraphrase at its face value, I therefore believe in the existence of entities that might be called "ways things could have been." I prefer to call them "possible worlds."

But, while these remarks do indeed give us some reason to believe in possible worlds generally, they do not yet support belief in the *concrete* possible worlds that Lewis believed in. His answer to the question of why we should believe that possible worlds are concrete is, quite simply, that abstract possible worlds cannot do all of the theoretical work that he thinks concrete worlds can do. Most importantly, abstract worlds either fail to represent all of the possibilities or else leave it a complete mystery as to how they represent any possibilities at all.

Before examining this objection to abstract worlds, let us briefly sketch the way in which Lewis thinks concrete worlds represent alternative possibilities. Remember that we are understanding "$\Diamond p$" to mean *there is a possible world in which p is true*. Where *p* is a sentence like "light travels faster than 186,000 miles per second" or "donkeys fly", it is easy to see how another concrete world might "represent" such possibilities. The possibility of donkeys flying is represented in another world if, and only if, the world contains donkeys and, lo and behold, the donkeys fly. What is harder to see is how other concrete worlds might represent possibilities for particular objects—*de re possibilities*, that is. So, for example, let *p* be the sentence, "Michael Rea is an Olympic archer". It is indeed possible that I be an Olympic archer. But the problem is that there is no concrete world other than this one that has me as a part. Worlds, on Lewis's view, do not share parts in common. (He gives various reasons for this. One simple reason is that, if they did share parts in common, then overlapping worlds would be spatiotemporally related to one another—which would make them parts of the *same* world.) And if no possible world other than this one has me as a part, then it is hard to see how another possible world could represent *me* as being an Olympic archer.

Lewis's solution to this problem is called *counterpart theory*. According to Lewis, what it is for an object in this world to exist "according to" another possible world is for it to have a counterpart in that world. Which things in other worlds count as an object's

counterparts is determined by similarity relations that hold between the object and things in other worlds. Our counterparts, then, are our representatives, or stand-ins in other worlds; and worlds represent possibilities for us by including counterparts of us who do things differently than we did. Thus, to say that I could have been an Olympic archer is, on this view, to say that there is another possible world in which a counterpart of me—a man distinct from but relevantly similar to me (by virtue of his history, intrinsic features, and so on)—*is* an Olympic archer.

More generally, counterpart theory affirms the following claims:

- *x* has a property F according to a world *w* if, and only if, *x* has a counterpart in *w* that has F
- *x* has the property of *being possibly F* if, and only if, there is a world according to which *x* has F
- *x* has the property *being necessarily (or essentially) F* if, and only if, *x* has F according to every world according to which it exists at all.

Note, however, that things in other worlds do not count as your counterparts simply by resembling you in some way or other. If they did do so, everything in every world would automatically be your counterpart, since everything in every world is similar to you in some way or other. You would have counterparts that are horses, rocks, atoms, electrons, and so on *ad infinitum*; and it would follow from this, on counterpart theory, that you have modal properties like *being possibly an electron* and *being possibly a rock*. So counterpart theory includes two further theses: (a) not every similarity relation is a counterpart relation, and (b) *which* similarity relations are counterpart relations is determined by context.

These two further theses are best understood in light of an example, one which we will discuss in more detail in Chapter 8. Consider a clay statue of Goliath (call it "Goliath"). In the region occupied by Goliath, we have a statue; we also have a lump of clay (call it "Lumpl"). Lumpl is just the lump of clay that Goliath is made of, the lump that *constitutes* Goliath. Being the lump of which Goliath is made, Lumpl occupies *exactly* the same region of space that Goliath does, and has all of the same material parts as Goliath. It is surely tempting to say

that Lumpl is just *the same thing* as Goliath. But notice: Goliath could not have been shaped as a ball; Lumpl could have been shaped as a ball. Thus, they have different modal properties—not everything that is possible for one is possible for the other. So, do we in fact have two things in the same place at the same time, instead of just one?

Lewis says no, and he explains our different judgments about the modal properties of Lumpl/Goliath as follows. Suppose we just point at the object and ask, "Could that thing have been shaped as ball?" There is, perhaps, not enough context to the question to enable us to make a judgment. But in a context where it is known that "Lumpl" is supposed to refer to a lump and "Goliath" is supposed to refer to a statue, the names "Lumpl" and "Goliath evoke different counterpart relations. That is, they give us different cues as to what ought to count as "relevantly similar" to the thing here in front of us. When we say that Goliath could not have been shaped as a ball, what we mean is that *this object has no statue-counterparts in other worlds that are ball-shaped*. When we say that Lumpl could have been shaped as a ball, what we mean is that *this object has a lump-counterpart in some other world that* is *ball-shaped*. Interestingly, counterpart theory also gives us a way of making sense of contingent identity claims. Lumpl is identical to Goliath; but counterpart theory, if correct, allows us to say truthfully that Lumpl could have been distinct from Goliath. When we say this, what we mean, according to Lewis, is that *there is a possible world in which either the object (Lumpl/Goliath) has a lump-counterpart that is not identical to any statue-counterpart of the object, or it has a statue-counterpart that is not identical to any lump-counterpart*.

What should we think of Lewis's theory? Perhaps the most common "objection" against the view that possible worlds are concrete is what Lewis has famously referred to as "the incredulous stare". People think that it is very strange to believe in a wide variety of other universes for the sorts of philosophical reasons Lewis gives for believing in them. Lewis has replied to this objection, however, by noting that "oddity is not falsity" and then by showing that belief in concrete possible worlds has a great deal of theoretical usefulness. For example, in addition to the advantages we have highlighted in the present chapter, we have seen that belief in other concrete possible worlds enables Lewis to provide a theory of properties (as sets of their instances) and of propositions (as sets of worlds). Lewis has also

invoked possible worlds to help build theories about counterfactuals (claims about what *would* or *might* have been the case had matters gone differently) and about mental content. Some of this work could be done by abstract worlds, but not all of it can be, and even the work that could be done by abstract worlds would have to be done differently.

Counterpart theory has also been a source of objection to Lewis's conception of possible worlds. One of the main worries is that our counterparts are just *other individuals* living in other universes, so it is hard to see what their activities have to do with possibilities for us. When I say that I could have been an Olympic archer (if I had a different set of skills and a different background and so on), I mean that *I* could have done this—the possibility in view is one that involves *me*, not some other guy who strongly resembles me. Lewis's response to this objection has been to insist that facts about our counterparts do represent possibilities that involve us. The *possibilities* involve us; the counterparts simply *represent* those possibilities. Still, many have found it hard to swallow the idea that what our counterparts in fact do is at all relevant to what we ourselves could have done.

Finally, opponents of Lewis's view have pressed the following more general objection as well. Not only do facts about our counterparts seem irrelevant to what is possible for us, but facts about Lewis's alternative possible worlds in general seem to have nothing to do with possibility and necessity. It seems that, had Lewis's concrete worlds not existed, it still would have been possible for things to be otherwise. I could have been an Olympic archer regardless of whether there is a concrete world in which some guy very much like me happens to be an Olympic archer. Indeed, the very fact that we can entertain the *possibility* in which no concrete worlds (except our own) exist ought to raise our suspicions. If all of the possibilities that exist are in fact represented by the concrete worlds, then it is *impossible* that there be no concrete worlds, since there is no concrete world that represents such a possibility. But it does seem possible that there be no concrete worlds; and it also seems possible that there not be enough concrete worlds to represent absolutely all of the possibilities. Let us turn, then, to the alternative: abstract possible worlds.

ABSTRACT POSSIBLE WORLDS

Whereas Lewis thinks of worlds as complete concrete universe-like objects, Alvin Plantinga, among others, takes worlds to be abstract states of affairs. We are well-acquainted with simple states of affairs: *Sally's sitting on the park bench*, for example. But there are also complex states of affairs corresponding to long and complex events that may or may not actually have happened: e.g., the French Revolution, or the coronation of the first king of the United States of America. Now imagine a state of affairs that comprises an entire possible history of the world, plus all of the metaphysical truths, plus all of the facts about mathematical objects and laws of nature and so on. Such a state of affairs would be a possible world.

We can express this idea more precisely as follows. Let us say that a state of affairs S is *maximal* if, and only if, it has the following feature: For *every* state of affairs $S^\star$, S either *includes* $S^\star$ or *precludes* $S^\star$. One state of affairs includes another if, and only if, the first can obtain only if the second also obtains. One state of affairs precludes another if, and only if, the two cannot both obtain. So a maximal state of affairs will either include or preclude *my being an Olympic archer*; it will either include or preclude *your reading this textbook*; and so on. A maximal state of affairs that is consistent (i.e., it does not include two states of affairs that preclude one another) is an abstract possible world.

According to Plantinga, a proposition p is true in a possible world w if, and only if, p would have been true had w been actual. Moreover, an individual exists in a world w if, and only if, the individual would have existed had w been actual. Likewise, an individual x has a property F in a world w if, and only if, x would have existed and would have been F had w been actual. And, again, "$\Diamond p$" means that there is a possible world in which p is true, and "$\Box p$" means that p is true in every possible world.

Earlier I said that one concern about abstract worlds pertains to how they manage to represent alternative possibilities. As we have seen, concrete worlds represent either directly or by way of counterparts. According to Lewis, there are only three possibilities for how an abstract world might represent something: It could represent in the way that sentences do, or it could represent in the way that pictures

do, or it could represent "magically" (i.e., in some other way that we can't really explain). Lewis has interesting objections against the idea that abstract worlds might be like pictures, but I'll set that whole discussion aside because the pictorial understanding of how worlds represent is, at best, a minority view. So the two options I'll discuss are these: Worlds are like sentences, or worlds represent by magic.

What would it mean for worlds to be like sentences? Sentences are made up of words, and the words pick out various bits of reality. So, for example, names pick out individuals; predicates pick out properties; and when you put them together, the sentence represents the individual as having the property. Different combinations of words, furthermore, generate different sentences and, therefore, different representations of reality. Maybe states of affairs are like this too: maybe, like structured propositions, they have parts that "mean" or "refer" to various bits of reality, and maybe different states of affairs are generated by different recombinations of those parts. (This is a very natural view for those who identify propositions with states of affairs.) If that is right, then possible worlds, on this view, will represent reality in the way that sentences do.

Lewis's main objection against this way of understanding possible worlds is that it is insufficient to represent all of the possibilities that there are. Words of real languages like English and German can only refer to objects that actually exist. We can't name merely possible objects or merely possible properties because we haven't the means to connect our words to those things. We can describe them if they are constructed out of properties or objects with which we are acquainted. For example, we can describe the property of being a gold mountain because that property is built up out of *being a mountain* and *being gold*, both of which are properties with which we are acquainted. But merely possible *fundamental* properties—properties that are in no way built out of properties we are acquainted with— will not be describable in this way. So if states of affairs represent the world in just the way that real-language sentences do, then they will be subject to the same limitations: they won't, indeed can't, have any parts that represent fundamental properties or particular objects that don't exist here in the actual world. This is the problem of *aliens*. If worlds represent in the way that linguistic items do, they haven't

the resources to represent certain kinds of alien properties or alien individuals.

Of course, one could insist that states of affairs aren't sentence-like, and that some of their parts somehow do manage to represent alien properties and alien individuals. But to say this is simply to move to the other option: the "magical" view. According to the magical view, there is no story to be told about *how* particular states of affairs represent the things that they do. So, for example, consider the state of affairs *my being a philosopher*. That state of affairs represents the real-world fact that I am a philosopher. But, on this view, we cannot explain *how* it represents this by saying (for instance) that it has parts—one corresponding to the word "philosopher", another corresponding to the name "Michael Rea", etc.—that have been assembled in just the right way to correspond to the fact that I am a philosopher. Nor can we point to some particular intrinsic feature of the state of affairs that explains why it represents the fact that I am a philosopher rather than (say) the fact that Notre Dame is a university. On this view, the state of affairs in question just *does* represent the fact that I am a philosopher, and that is all there is to be said about it. Lewis's objection, in short, is that the magical view makes representation wholly mysterious.

So, according to Lewis, believers in abstract worlds must admit either that abstract worlds cannot represent all of the possibilities there are (the problem of aliens) or that they have no idea what the representation relation amounts to (the problem of mystery). The first horn of the dilemma seems clearly unacceptable. To admit that your theory of possible worlds lacks the resources to represent all of the possibilities is to admit that it is a failure as a theory of worlds. The second horn of the dilemma, however, is not so obviously problematic. Not everything can be explained; perhaps here is a place where we must rest content with mystery. At any rate, the problem of mystery doesn't point to any outright incoherence in the view that worlds are abstract. Nor does it show that the view lacks the resources to do what it claims to be able to do. Still, what Lewis will claim is that the problem of mystery counts as a theoretical disadvantage. It is a coin on the *concrete worlds* side of the scale. The question, then, is whether the overall balance of coins tips the scale in the direction of Lewis's view or in the direction of Plantinga's.

There is one other issue to discuss in connection with the abstract-worlds theory. It is a problem that is in some ways connected with the problem of aliens. Consider a possible world in which Socrates does not exist. Surely from the point of view of that world it is *possible* that Socrates exist. Thus, from the point of view of that world, Socrates is an alien individual. As we have seen, there is, then, a question to be raised about how the world represents the possibility that Socrates exists. But there is another question as well. How does the world represent Socrates himself as *not existing*? This is, of course, just another form of the problem of nonexistent objects. We have already seen a few strategies for dealing with this sort of problem in Chapter 2; but it is worth revisiting the issue briefly so as to make it clear how the problem connects up with our discussion of possible worlds.

Another way of putting our question is: How could it be true in a world *w* that Socrates does not exist? The question poses a problem because, intuitively, a proposition *p* is true in a world *w* if, and only if, *p* would have been true had *w* been actual; and it is initially hard to see how a proposition like *Socrates does not exist* could possibly be true. For it seems that the truth condition for the proposition that Socrates does not exist would have to be something like the following:

TC1: The proposition that Socrates does not exist is true if, and only if, Socrates exemplifies nonexistence.

TC1, in turn, suggests the following more general condition:

TC2: For any *x*, the proposition that *x* does not exist is true in a world *w* if, and only if, *x* would have exemplified nonexistence had *w* been actual.

But TC1 and TC2 are hard pills to swallow. As we saw in Chapter 2, there is something deeply puzzling about the idea of something exemplifying nonexistence. For a *thing x* can exemplify nonexistence only if there *is* such a thing as *x*, despite the fact that *x* does not exist. In other words, TC1 and TC2 presuppose that being and existence come apart, which seems impossible.

Significantly, Lewis's theory of worlds does not face this problem. A world *w* in which Socrates does not exist is one which neither has him as a part nor includes a counterpart of Socrates. Is it *true* in *w* that Socrates does not exist? Yes; but that truth is not problematic for Lewis. For what is true, strictly speaking, is just that Socrates does not exist *in w*; it is not true, strictly speaking, that Socrates does not exist. The reason is that, although Socrates is not part of *w* and so does not exist in *w*, it remains the case that, from the point of view of *w*, Socrates does exist, but in some other concrete world.

The problem arises, then, primarily for believers in abstract worlds. One solution, of course, is to embrace TC2. Those who take this solution maintain that, in addition to all of the things that actually exist, there are also *merely possible* things that do not actually exist. This is a theory known as "possibilism". Possibilism is different from views (mentioned in Chapter 2) that maintain that even *impossible* objects, like round squares, have being. This is because TC2 does not by itself imply that there are or could have been such things. But it is obviously related to those views. The opposing view, according to which everything that has being exists, is known as "actualism". I have nothing further to say about possibilism here except that I cannot imagine how it could be true except by abandoning a conception of possible worlds as abstract, and embracing, as Lewis does, the idea that worlds are concrete and that "merely possible things" are just concrete inhabitants of worlds spatiotemporally disconnected from the actual world.

How, then, might an actualist solve the problem?

There are at least two strategies, both of which were discussed in Chapter 2. The first is to insist that everything exists necessarily—so, in other words, to deny that there are possible worlds in which Socrates (or anything else) fails to exist. This solution sounds incredible; but, as we saw in Chapter 2, it has been defended in recent years by Timothy Williamson.

The other strategy is to say that, in worlds where Socrates does not exist, the proposition that Socrates does not exist is true because properties like *socrateity* or *being identical to Socrates* are unexemplified. Properties like these are *individual essences* of Socrates. A property *P*

is an individual essence of an object *x* if, and only if, (a) *x* cannot exist without having *P* and (b) necessarily, nothing other than *x* has *P*. Unlike Socrates himself, essences of Socrates exist (according to proponents of this solution) in *every* possible world. Thus, there is no problem invoking them to account for the truth of propositions like *Socrates does not exist*. Thus, in general, this solution maintains that a proposition of the form *x does not exist* is true in a world *w* if, and only if, every individual essence of *x* would have been unexemplified had *w* been actual.

As we saw in Chapter 2, the trouble with this strategy is that it is unclear how individual essences—which, by definition, are not *multiply* exemplifiable—can possibly count as properties; and it is also unclear how they can be what they are without bearing some relationship to the objects of which they are essences. I have nothing more to say here about the first of these two problems; but, with our discussion of possible worlds now on the table, there is a bit more to say about the second one. The second problem gets traction from the idea that a property like *socrateity* somehow *represents* Socrates (or is in some relevantly similar way *about* Socrates), and it is hard to see how it could do this without standing in some relationship to Socrates. But at this juncture it seems that the proponent of abstract worlds *could* respond to the problem by saying that, in worlds in which Socrates does not exist, *socrateity* does indeed exist and *represent* Socrates, but it does so by "magic", not by way of any particular relationship to Socrates. This is difficult claim to accept, of course; but at least it assimilates this problem to a problem most believers in abstract worlds already face—namely, the problem of abstract representation in general.

FURTHER READING

On the topic of parallel universes, which was mentioned and then set aside at the beginning of this chapter, a good starting place is Max Tegmark's "Parallel Universes", reprinted in *Arguing about Metaphysics.* For discussion of parallel universes in connection with the Fine Tuning Argument, see John Leslie, "World Ensemble, or Design", reprinted in *Arguing about Metaphysics,* Robin Collins, "A Scientific Argument for the Existence of God", in Michael Murray (ed.), *Reason for the Hope Within* (Grand Rapids, MI: Eerdmans, 1999), 47–75 and Stephen

Barr, *Modern Physics and Ancient Faith* (Notre Dame, IN: University of Notre Dame Press, 2006).

On the topic of possible worlds, the best entry points are John Divers, *Possible Worlds* (London: Routledge, 2002), and the superb introduction to Michael Loux (ed.), *The Possible and the Actual* (Ithaca, NY: Cornell University Press, 1979), to which my own discussion here is heavily indebted. Also in *The Possible and the Actual* are quite a few of the most important articles from the discussions of modality taking place in the 1970s in the wake of the publication of Kripke's 1963 article, "Semantical Considerations on Modal Logic", *Acta Philosophica Fennica* 16 (1963): 83–94.

I cited David Lewis's book, *Counterfactuals* (Oxford: Basil Blackwell, 1973); but the most important place to look for his views about modality is his *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986). I have also found Philip Bricker, "David Lewis: On the Plurality of Worlds", in *Central Works of Philosophy,* vol. 5, *The Twentieth Century: Quine and After* (Chesham: Acumen Publishing, 2006) to be very useful; and, as mentioned at the end of Chapter 2, Daniel Nolan's *David Lewis* is an excellent overall guide to the philosophy of David Lewis.

Alvin Plantinga's views about modality can be found in *The Nature of Necessity* (Oxford: Oxford University Press, 1974) and in the essays collected in Alvin Plantinga, *Essays in the Metaphysics of Modality*, edited by Matthew Davidson (Oxford: Oxford University Press, 2003). *Arguing about Metaphysics* reprints an excerpt from Lewis's *On the Plurality of Worlds*, as well as an article by Plantinga— "Two Concepts of Modality"—that highlights the differences between his own views and those of David Lewis.

5

TIME

In an oft quoted passage, St Augustine remarks, "So what is time? If no one asks me, I know; if I want to explain it to someone who asks me, I do not know." We know time experientially. There is probably no more fundamental or constant aspect of our experience than our sense of time and its passage. Knowing it experientially is one thing. Understanding it is another.

Time slips away, and the future has yet to come. So it is natural to think that there is, quite literally, nothing but the present. But the present is so incredibly "thin" that it seems hardly big enough to contain all the richness of thought and experience and activity that occupies our minds and lives at any given moment. Could it really be that all there ever is to a high school prom, or a college football game, or a bloody war is the thinnest slice of instantaneous action? Is there nothing more to time itself than an eternal succession of tiny slivers of temporality? Or are past and future more like distant places that not only exist but are even accessible, if only we had the right technology or a proper understanding of physics?

The first step toward understanding time is to say something about the nature of *times*, and about what the terms "past", "present", and "future" refer to. The next step is to look at what reasons we might have for and against the claim that past and future times (or, as I shall

prefer to put it: past and future objects and events) do not exist. Discussing this question will, in turn, raise questions about the *passage* of time. What, exactly, does it mean to say that time passes? *Can* time pass? These are the issues in focus in the present chapter.

TIMES

Says Qoheleth, "For everything there is a season, and a time for every matter under heaven: a time to be born, and a time to die" (Ecclesiastes 3: 1–2, NRSV translation). Ordinary speech, and much poetry, is replete with reference to times. What are these things?

There are three main ways of thinking about the nature of times. On one way of thinking, times are events. Some people think of events as abstract entities. For purposes here, however, we shall assume they are concrete. We can watch them happen, participate in them, cause them, and be affected by them. They are wars and weddings, games and celebrations, births and deaths. They exist precisely when they happen. Some exist only for an instant; others last through hours, days, years, even centuries. So now consider some arbitrary event—the beginning of your own existence, for example. Let us suppose that this event happens instantaneously, rather than gradually, over some duration. Now consider every event that was simultaneous with that one (in some particular **frame of reference**). Now consider the grand event that is the **sum** of all of those smaller ones. That grand event is a time. (I assume here that for any two events, there is exactly one further event that is their sum. So, for example, in addition to the football game down at the stadium and the tailgate barbecue in the parking lot, there is also an event consisting of *both* the game and the barbecue.)

We can represent this view graphically as follows. Consider Figure 1. The horizontal axis represents space, the vertical axis represents time, the dots represent particular instantaneous events happening at each time, and the open space on the graph represents spacetime itself. The view that times are events would identify the time *t1* with the sum of *everything* that is happening at the *t1* coordinate. So *t1* will include *a* and *b* as parts. If, as many physicists believe, there is stuff going on even in what we ordinarily think of as "empty" space, we could shade in all of the open space on the graph and *t1* would include those
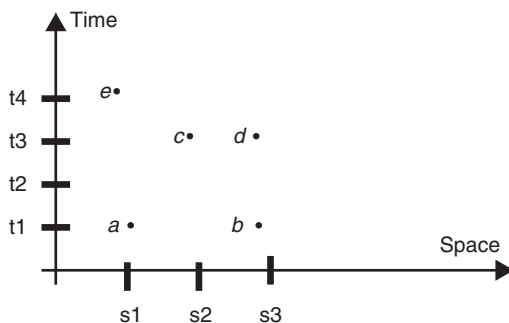
Figure 1

events as well. It will also include other events that are only implicitly represented on the graph, like the event of *spaces* s1 *and* s2 *being empty*, and a'*s occurring at such-and-so distance from* b.

Notice that if times are events, it follows that past and future times exist only if past and future events exist. Indeed, it also follows that if the time of your birth exists, then your birth itself exists (since your birth is a part of that time). Likewise, if future times exist then all of the events that happen at those times exist as well.

Another way of thinking about times is to think of them as abstract states of affairs. Corresponding to every event is a state of affairs that obtains if, and only if, the event is happening. The occurrence of the event, we might say, *is* the obtaining of the state of affairs. (But again, the event, unlike the state of affairs, exists *only* when the event is happening.) So too, then, with times: Corresponding to each grand event that we have been calling a "concrete time" there will be an abstract state of affairs. We might choose to identify the time with the state of affairs rather than with the event.

As we have seen, it is common to think that states of affairs exist necessarily if they exist at all. Given this, treating times as states of affairs settles the question whether past and future times exist (of course they do). But it still leaves open the more interesting question of whether past and future *events* exist.

Finally, one might think of times as *locations*, like points or planes in space. Consider our graph again. The view that times are locations would say one of two things. (a) Times are not to be identified

with particular events, or states of affairs, or any other existing thing. Rather, existing *at a time* is a matter of standing in certain "earlier than" and "later than" relationships to other things. So, for example, existing *at t2* (in Figure 1) is a matter of being a little earlier than *c* and *d,* a little later than *a* and *b*, and so on. Or (b) there is something in reality that corresponds to the blank paper on which the graph is drawn—a collection of empty "places" where events or the substance of spacetime itself can be inserted. A *time* is a particular region within that collection of empty places, a region like the empty lines demarcated by the labels "*t1"*, "*t2",* etc., on the graph.

I will not advocate here for any particular view about the nature of times. But I do want to note that what one thinks about the nature of *time itself* will depend upon which of the aforementioned views one endorses. Likewise, what one thinks about what "past", "present", and "future" refer to will also depend upon what one thinks about times. If times are concrete events, then the past comprises just those grand events that happened prior to the present moment; and time itself is the sum of whatever grand events exist. If times are abstract states of affairs, then time will also be some sort of abstract thing, perhaps a total world history, The past, on this view, will be a partial world history: the state of affairs which includes everything that has ever happened up to the present moment. Likewise for the future.

For the remainder of this book, I will assume that times are concrete events.

PAST, PRESENT, AND FUTURE

Let us now consider the ontological status of non-present objects and events. Are there such things? On the one hand, it is hard to believe that events in the past and the future—your birth, your death, the age of the dinosaurs—are somehow "just as real" as events in the present. Where, after all, would such things be? On the other hand, it seems obvious that we can *remember*, *talk about*, *study*, and stand in all manner of other relations to merely past objects and events. (Some people remember the Vietnam War, for example; others merely study it; I am presently writing a sentence about it.) But, as noted in Chapter 2, it is hard to see how one can stand in relations to things that do not exist.

Some philosophers believe that there are no merely past or merely future things. They endorse *presentism*, the view that it always has been and always will be the case that whatever exists *presently* exists. Presentism might seem to be a truism. After all, "exists" is a present tense verb, so *of course* whatever exists presently exists. Merely past things exist*ed*; things lying in the distant future *will* exist. To say that they *do* exist, then, might seem to reflect a failure to understand tense. But to think in this way is to miss the point of the debate between presentists and their rivals. Notice that in explaining why presentism might seem to be a truism, I made free use of expressions like "merely past things" and "things lying in the distant future". From the presentist's point of view, these are not referring expressions. Nothing lies in the distant future, and nothing is merely past. Strictly speaking, then, it is not true (according to a presentist) that merely past things existed, since there is not, never has been, and never will be anything referred to by the description "merely past thing". This, then, is something that presentists and their opponents can disagree about, which goes to show that presentism is not a truism after all.

The alternative to presentism is *four-dimensionalism*, which I will characterize simply as the thesis that presentism is false. Four-dimensionalism is currently the majority view among those writing in the field, in no small part because it is the view that many think fits best with our most successful theories in physics. In particular, most philosophers endorse a version of four-dimensionalism known as *eternalism*, the view that everything that ever did exist or ever will exist *does* exist.

"Scientific people," says the Time Traveler in H. G. Wells's *The Time Machine*, "know very well that Time is only a kind of Space." In the story, this remark comes toward the end of a brief parlor lecture wherein the Time Traveler argues that reality is extended in four dimensions, not three, and that the so-called temporal dimension is not fundamentally different from the so-called spatial dimensions. The idea, then, is that the universe is *not* a three-dimensional spatial volume that "moves" through time from one moment to the next; rather it is a four-dimensionally extended block whose total volume *includes* its entire history. This is not to say that eternalism is inconsistent with the idea that time somehow passes. If the idea that time

passes is itself consistent, then there is no special reason to think that eternalism precludes the passage of time. Rather, the point is that eternalism implies that the universe as a whole doesn't *move through* time; instead, the universe itself has a temporal dimension just as it has spatial dimensions. (This is not to say, however, that, it has exactly one "objective" temporal dimension. Relativity theory is typically understood to imply that there are many equally good ways of dividing spacetime into spatial and temporal dimensions.)

Although eternalism is the most popular version of four-dimensionalism, it is not the only version. Among the other versions are views according to which our universe is a *growing block* or a *shrinking block*. According to the former, only the past and the present exist, and the block of reality continues to grow as new times come into existence and existing times change from being present to being past. According to the latter, only the present and the future exist, and the block of reality continues to shrink as times cease to exist. There are still other varieties as well, but for the remainder of this section we shall leave them all aside. Going forward, when I talk about considerations for and against four-dimensionalism, it shall be eternalism that is primarily in view, though readers should bear in mind that many of these considerations will apply equally to the other varieties of four-dimensionalism.

As we have seen, the Time Traveler cited *science* in support of his claim that "Time is only a kind of Space". Interestingly, *The Time Machine* was written in 1895, ten years before the special theory of relativity revolutionized our understanding of space and time. Nevertheless, Einstein's theory lends support to the thesis that space and time as we know them in experience are in fact mere appearances of a more fundamental reality—namely, spacetime. The *locus classicus* for the claim that special relativity supports this latter thesis is H. Minkowski's "Space and Time", which includes the famous remark that "Henceforth, space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality." General relativity, proposed in 1916, likewise lends support to the thesis that there is no fundamental difference between space and time. Unsurprisingly, then, it is standard for descriptions of the geometry of spacetime to treat past and future events (within a frame of reference) as relevantly

like distant places—existing elsewhere, but no less real than events happening here and now.

Some have made stronger claims on behalf of eternalism, arguing that its denial is outright inconsistent with our best scientific theories. These stronger claims, however, are generally regarded as having been discredited. Nevertheless, there are other considerations that speak in favor of eternalism. I shall focus on two.

The first, which I have already mentioned, is the so-called "problem of cross-time relations". It is hard to see how we can refer to and stand in other relations with objects and persons in the past and future unless such things exist. How can I remember my wedding if my wedding does not exist and, therefore, cannot be the object of any memory? How can I admire the bird who struggled so fiercely to escape the clutches of my cat when the bird no longer exists to be an object of my admiration? How can I sensibly think that my present acts can have future consequences if I insist that nothing but *present* events are available to stand in the *is caused by* relation to my actions?

We can convert these rhetorical questions into straightforward arguments for the existence of past and future objects and events. Thus:

5.1   I remember my wedding day, and I admire the bird who struggled with my cat.
5.2   Therefore: There is an x such that I remember x and x is identical to my wedding day; and there is a y such that I admire y and y is identical to the bird who struggled with my cat.
5.3   Therefore: My wedding day exists, and the bird who struggled with my cat exists.

From 5.3 and the fact that neither my wedding day nor the bird *now* exist, it follows that at least one merely past event and at least one merely past object exist. Likewise, we can argue for the existence of future events as follows:

5.4   My present acts are causes of events that are not happening now but will happen in the future.
5.5   Therefore: There are events x and y such that x is a present act of mine, y is an event that is not happening now but will happen in the future, and x is a cause of y.

5.6   Therefore: An event that is not happening now but will hap-
         pen in the future exists.

From 5.6 it follows that there exists at least one merely future event.

One lesson from the first two chapters of this book is that we
should be very suspicious of "cheap arguments" like these. True
enough, they are straightforward. But, as we have seen, there are argu-
ments like these to be had for the existence of all manner of things
that we do not want to believe in. The challenge, then, for those who
do not want to believe in past times is to explain why the first prem-
ise in each argument is false, or why the second does not follow from
the first, or why the third does not follow from the second. I will take
these strategies in reverse order, and I will restrict my focus to the
first of the two arguments.

To deny the inference from 5.2 to 5.3 is to say that "there is some-
thing identical to my wedding day" does not imply that my wedding
day exists, and that "there is something identical to the bird who
struggled with my cat" does not imply that said bird exists. We have
already explored this sort of move in Chapter 2, and there is nothing
further to say about it here.

The better option, it seems to me, is to deny that 5.2 follows from
5.1. To do this is to say, in effect, that memory and admiration do not
require *objects*. Perhaps, what it is to *remember my wedding day* is for
me to have a certain kind of mental imagery that stands at the end
of a particular causal chain which began on my wedding day. Perhaps
what it is for me to admire the bird who struggled with my cat is for
me to have feelings of admiration that stand at the end of a causal
chain which began with my initial observations of the bird. If all of
this is correct, then remembering my wedding day does not involve
standing in relations to my wedding day, and so it doesn't require
the existence of my wedding day. All it requires is that my wedding
day exist*ed*, and that it gave rise to a causal sequence culminating in
particular mental imagery that I now have. Likewise for admiration
of the bird.

The problem with this reply is that it relies on the possibility of
diachronic (cross-time) causation—i.e., causation between things
existing at two different times. For the reply to work, there needs to
be a causal *chain* moving forward in time from my wedding day to

my present mental imagery, or from the event of the bird's struggle to my present feelings of admiration. But if such causal chains can exist only if past events exist, then the reply we are considering is useless.

The third alternative, then, is to deny the data: contrary to appearances, I *do not remember my wedding day*; and, contrary to appearances, I *do not admire the bird*. In light of what has just been said about the second reply, it should be clear that, in the end, proponents of this reply will also have to deny the possibility of diachronic causal relations.

The other primary (non-scientific) motivation for eternalism comes from reflection on the apparent impossibility of temporal passage. Reflection on the apparent *experience* of temporal passage provides one of the primary motivations for rejecting a "static" eternalist theory of time in favor of "dynamic" rival theories, such as presentism and the growing block theory. (Dynamic eternalism is also an option, although not a popular one.)

TEMPORAL PASSAGE

In his famous article, "The Myth of Passage", D. C. Williams observes that the "most substantial and incorrigible" motivation for believing in the passage of time is the felt experience of temporal passage:

> It is simply that we *find* passage, that we are immediately and poignantly involved in the jerk and whoosh of process, the felt flow of one movement into the next. Here is the focus of being. Here is the shore whence the youngster watches the golden mornings swing toward him like serried bright breakers from the ocean of the future. Here is the flood on which the oldster wakes in the night to shudder at its swollen black torrent cascading him into the abyss.

Williams himself does not believe that time passes, so, strictly speaking, he should not concede that we *find* passage. For we can find only that which is real. Nor likewise should we open debate about the reality of temporal passage with the claim that we *experience* passage. For we can experience only that which is real. Instead, we should follow Laurie Paul's more careful mode of expression and ask not what we ought to believe in light of the experience *of* passage, but rather what we ought to believe in light of the experience *as of* passage, thus leaving open the question whether our experience is veridical.

Static eternalism is the thesis that eternalism is true and time does not pass. *Dynamic eternalism* is the thesis that eternalism is true but time does pass—presentness moves along the temporal dimension sort of like a spotlight moving along a wall. The question I now want to ask is whether our experience as of passage provides strong evidence that time passes. If it does, then it provides strong evidence against static eternalism.

Philosophical arguments against the reality of temporal passage imply that our experiences as of passage are illusory. It may be hard to give these arguments much credence if we have no independent reason for thinking that we might be susceptible to the illusion of passage, and no explanation of how such an illusion might arise. Under such circumstances, one quite naturally weighs the evidence coming from constant experience more heavily than the evidence coming from philosophical intuition. However, Laurie Paul observes that it is already a well-known empirical fact that human beings are susceptible to illusions of motion, and she argues that the illusion of passage might arise in a way very similar to the way in which some of these other illusions arise. If she is right, then the hypothesis that our temporal experience is illusory is a live option. If it is a live option, then the philosophical arguments against the reality of passage should carry more weight.

There are various well-known perceptual illusions in which subjects have experiences as of motion where in fact there is none (or, at any rate, none of the sort they seem to be experiencing). For example, in a string of lights where each light is successively turned on and then off, we will have an experience as of the light racing across the string. On a computer monitor, if different dots are rapidly shown in succession in different places along a straight line, we will have an experience as of a dot moving across the string. Paul also notes that films, time-lapse photography, and flip-books also provide us with illusions of motion. These cases alone suffice to establish the claim that we are *susceptible* to illusions of motion. According to Paul, empirical research on similar cases helps to provide an explanation of how the illusions might arise—an explanation which, she argues, sheds light on how we might have illusory experience as of temporal passage.

According to Paul, empirical research indicates that what happens in these cases is something like the following. *First*, the brain processes

our experiences of the relevant static images (successive flashing dots, or the pictures on the pages of a flip-book); *then*, afterwards, it mistakenly represents them as a single *continuously moving* image rather than as multiple static images occurring in sequence. If that is right, then the experience as of motion neither involves nor requires the continuous perception of motion. Instead, motion is represented "all at once" in a static brain state that is itself the product of prior sensory processing. So the illusion of motion can arise in the cases just described precisely because the continuous perception of motion is not required for the representation of motion.

Paul's suggestion, then, is that the illusion of temporal passage might arise in a similar way. Suppose, as many eternalists would maintain, that at every moment *m* of your brain's existence there exists a distinct "brain-stage", a part of your temporally extended brain that exists at, and only at, that single moment *m*. (The idea that there might be temporal stages of four-dimensionally extended material objects like brains will be taken up in more detail in Chapter 9.) Then what *might* take place is something like the following. First, a series of brain stages might together realize all of the various steps involved in *processing* sensory input from two discrete times; then another brain stage (or series of brain stages) might mistakenly represent that sensory input as being the product of a moving sequence of events which includes the "jerk and whoosh" of temporal passage.

This is hardly a complete or uncontroversial account of how the illusion of passage might arise. But the availability of this sort of empirically grounded sketch of an explanation makes it difficult simply to dismiss out of hand the idea that our temporal experience might be illusory. It is at this point, then, that philosophical arguments against the possibility of temporal passage come to have real bite.

The problem is that it is hard to see what temporal passage could possibly amount to. Words like *flow* and *passage* suggest a kind of movement. A very natural thought is that the passage of time is the movement of "presentness" along a timeline—the "moving spotlight" idea mentioned earlier. This is not the only way to think of temporal passage, of course. Indeed, only eternalists could think of it this way since only eternalists believe in anything remotely like a *timeline* along which presentness might *move*. But before introducing

the other main way of thinking about temporal passage, let us pause for a moment to reflect on this one.

There is a venerable argument, due to J. M. E. McTaggart, that purports to show that temporal passage construed in the way just described is impossible. To understand McTaggart's argument, it helps to have in hand some of his terminology. A series of events whose members have properties like *being past*, *being present*, and *being future* is called an **A-series**. Accordingly, the properties themselves are commonly called A-properties. A **B-series** is a series of events whose members are ordered by simultaneity and temporal priority (earlier-than and later-than) relations. The relations themselves are commonly called *B-relations*. Any A-series is also a B-series, but not vice versa.

McTaggart assumed—and most philosophers are willing to grant—that if time exists at all, then time is *at least* a B-series. The question, then, is whether time must also be an A-series. Accordingly, philosophers typically say that an **A-theory** of time is one according to which time, if it exists at all, is both an A-series and a B-series, and a **B-theory** of time is one according to which time, if it exists at all, is just a B-series. McTaggart was an A-theorist. He thought, and argued, that time *cannot* be just a B-series. However, he also thought that an A-series is impossible. His argument for that conclusion is his famous argument against the possibility of temporal passage. If the argument is sound and the A-theory is right, then the ultimate conclusion of his paper is established: Time is unreal.

Simplified a great deal, McTaggart's argument against temporal passage starts like this:

5.7   Suppose (for *reductio ad absurdum*) that an A-series exists.
5.8   If an A-series exists, then events change their A-properties.
5.9   If events change their A-properties, then one of the following is true about *every* event *e*:
    a   *e is* future and *will be* present, and then past.
    b   *e is* past and *was* future, and then present.
    c   *e is* present, *was* future, and *will be* past.

So far so good. But then McTaggart makes a surprising move: He suggests that if 5.9 is true, then *every* event has *every* A-property. Reconstructing the reasoning behind this move has proven to be

extraordinarily difficult; but my own view is that McTaggart gets to his conclusion by way of the following assumption:

5.10   For any object x and property ϕ, if x is, was, or will be ϕ, then x is ϕ.

If 5.10 is true, then (for example) if *e will be* past, *e is* past. So if an event *e* is present and will be past, then it is both present and past. If it was future and is present, then it is both present and future. So it follows from 5.8–5.10 that, if an A series exists, every event *e is* future, *is* present, and *is* past—i.e., every event has every A-property. Thus, the argument continues:

5.11   Therefore: If an A-series exists, every event has every A-property. (From 5.8–5.10)

Together with 5.7, 5.11 implies that every event has every A-property. But it is obvious that no event can have more than one A-property. So we reach a contradiction, thereby reducing 5.7 to absurdity. Since the supposition that an A-series exists leads to a contradiction, it cannot be true. An A-series, therefore, is impossible.

The crucial premise in the argument is 5.10. At first glance, the premise seems obviously false. Why think that, in general, if something *was* ϕ then it *is* ϕ? Beauty fades; so it hardly follows from the fact that someone *was* beautiful that they *are* beautiful. Things deteriorate; so it hardly follows from the fact that the building *was* sturdy that it *is* sturdy. Who in their right mind would affirm 5.10?

Answer: An eternalist. Indeed, the premise seems irresistible if one is an eternalist. According to the eternalist, every event that ever did exist or will exist does exist. Notice that the phrase "does exist" is used here in a tenseless way. Nobody would say that merely past or merely future events *presently* exist. But eternalists (and growing block theorists) still say that they exist. That makes sense only if "exist" can be used in an untensed way. Once we see this, it is much easier to appreciate why the eternalist would affirm 5.10.

For clarity's sake, let us in the remainder of this section signal important tenseless uses of words by formatting them in all-caps. Premise 5.10 should then be rewritten as follows:

5.10   For *any* object *x* and property φ, if *x* is, was, or will be φ, then *x* IS φ.

(Other premises will have to be rewritten as well; but let us restrict our focus to 5.10.) We can now understand the eternalist's reasoning as follows. Suppose *e* was present but is now past. Whatever did, does, or will exist EXISTS. So the event *e's being present* EXISTS. But, of course, that event EXISTS only if *e* IS present. So, given eternalism, if *e* was present, then *e* IS present. But nothing in this reasoning depends on peculiarities about *e*, presentness, or the past tense (as opposed to the future tense). So we can generalize: for *any* object *x* and property φ, if x is, was, or will be φ, then *x* IS φ. Hence 5.10 is true.

That is one reason for affirming 5.10. Here is another (related) reason, which does not appeal directly to eternalism but rather rests on a particular analysis of tense. McTaggart seems to think that to say that *x* will be φ is to say that there IS a moment of future time at which the event of *x's being* φ IS happening. Likewise with past tense claims. But if that is right, then if *e* will be past, *e* IS past; if *e* was future, *e* IS future; and so on.

How might we respond to these defenses of 5.10? One way is to suppose that the "moment of future time" at which *e's being past* occurs is a moment in some other, time-like series. Thus, we might say something like this: there are locations, T1 and T2, in another temporal series (call it "hypertime") such that *t1* is the present moment at T1 and *t2* is the present moment at T2. (See Figure 2.)
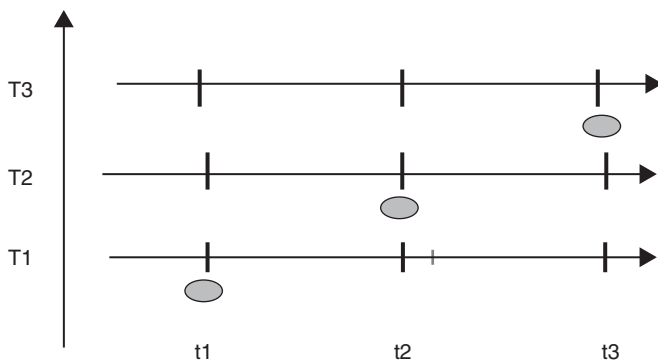

Figure 2

The basic idea here is that T1, T2, etc. are coordinates in *another* temporal series, and that the "timeline" of our universe exists in its entirety at each of those coordinates, differing at each coordinate only with respect to which events are present. In Figure 2, the dot represents where *presentness* is located at each hypertime. On this picture, then, the passage of *time* (i.e., passage within the *t-series*) consists in presentness changing its location on the t-series from one hypertime to the next. In Figure 2, for example, it seems true that, at T1, *t1* is present but will be past, whereas at T2, *t1* was present and is past.

This response helps us to see how one might accept McTaggart's analysis of tense without accepting 5.10; but it does so only at the cost of leading us into an infinite regress. To see why, consider this question: Is there such a thing as the *passage of hypertime*? Or is a B-theory true for hypertime? If we affirm a B-theory for hypertime, then it looks as if 5.10 will still be true. Look again at Figure 2. Suppose that time passes in the *t*-series, but not in the T-series; and suppose we are presently located at T1. Still, if the T-series is a B series, T2 and T3 EXIST, as does everything that HAPPENS at those times. So our problem remains. On the other hand, if we affirm an A-theory for hypertime, then won't we have to believe in a third time series, hyper-hypertime in order to make sense of passage in hypertime? An infinite regress looms, which would clearly be a disaster.

My own view is that, given the truth of four-dimensionalism, McTaggart's argument is decisive. If four-dimensionalism is true, time does not pass. One of McTaggart's contemporaries, however, famously disagreed. C. D. Broad, a growing block theorist, declared McTaggart's argument to be "a howler". Broad insisted that McTaggart failed to see that temporal passage consists in the *absolute generation of new events*. If he is right about the nature of temporal passage, then McTaggart has indeed made a mistake in analyzing future tense claims in the way suggested above. But by accepting the growing block theory, it looks as if Broad has solved, at most, only half of the problem. For the growing block theorist maintains that past events EXIST. And if they do, then a suitably modified version of McTaggart's argument will show (absurdly) that every event has *two* A-properties—being past and being present.

So I think that the objections to 5.10 thus far considered are failures. Astute readers will notice, however, that if 5.10 is true, it is not just change in A-properties that will be problematic, but *any change*

whatsoever. There is, then, a much more general problem lurking in the neighborhood, and four-dimensionalists will have to find a way to grapple with it.

Ultimately, then, I think that McTaggart's argument is decisive under the assumption that four-dimensionalism is true. What this means, is that dynamic versions of four-dimensionalism—dynamic eternalism with its moving spotlight theory of temporal passage, the growing block theory, the shrinking block theory, and so on—are unworkable. In light of this, it is easy to see why the *experience* of temporal passage would be taken as evidence in support of presentism. But taking it so presupposes that presentists have a workable reply to McTaggart's argument. Is this correct?

Yes, and no. Like Broad, the presentist will say that McTaggart has overlooked the possibility that temporal passage consists in the absolute generation of new events. Unlike Broad, the presentist will *deny* that there is any A-series of events. For presentists do not believe that there *ever* exists a temporally ordered *series* of events at all. Rather, temporal passage consists in the ongoing generation *and destruction* of times. So, interestingly, presentists can accept McTaggart's argument for the conclusion that *there is no A-series*, and they can insist that the whole problem arises out of treating temporal passage as change in the members of a series of events rather than as the ongoing generation and destruction of times.

Be that as it may, presentists are still generally considered A-theorists, and this for two reasons. First, on some ways of thinking about times and events, it does turn out that presentists believe in an A-series. For example, if times are abstract, then every time that ever did or will obtain does *exist*; and those times do change their A-properties. Second, even if the only A-property, strictly speaking, is *presentness*, it is not unreasonable to think that coming into and passing out of existence is a way for a time to "change" its A-property. And, clearly enough, so long as a presentist accepts 5.10, it will be easy to show that their view is inconsistent with temporal passage. For if presentism is true and time passes, at least one of the following claims must be true about some particular event *e*:

5.12   *e was* present, but *is* not present.
5.13   The proposition that *e* is present *was* true but *is* not true.
5.14   A sentence saying that *e* is present *was* true but is not true.

Any of these claims, however, in conjunction with 5.10 will imply that *e* IS present and that *e* IS not present, which is contradictory. The only non-superficial reply for a presentist to make to the argument, then, is to reject 5.10.

Rejecting 5.10 is a pretty obvious move for a presentist to make in any case. Why would they even consider accepting it? Consider the claim that *e was* present. Unlike the eternalist, the presentist does not believe that the event of *e's being present* EXISTS; so there is no reason for the presentist to think that *e* IS present.

Keep in mind, however, that whatever else a presentist says about McTaggart's argument, they are ultimately committed to the view that events, even times, are continually coming into and passing out of existence. But this is deeply puzzling. What could it possibly mean, given presentism, to say that some particular event, or time, *did* exist or soon *will not* exist? How can we understand the tense in these claims? There seem to be only two options. One is to accept McTaggart's analysis of tense. But we have already seen that this way lie problems. The other option is to say nothing, to insist that tense is sufficiently well-understood that no account of it is needed. This is not an incoherent position, but it leaves the phenomenon of temporal passage shrouded in mystery. Those who cannot abide the mystery will see in McTaggart's argument strong evidence for the conclusion that presentism is false and time does not pass.

FURTHER READING

The opening quotation from St Augustine's *Confessions* (11.14) comes from Arthur Hyman, James J. Walsh, and Thomas Williams (eds), *Philosophy in the Middle Ages: The Christian, Islamic, and Jewish Traditions*, 3rd ed. (Indiananoplis, IN: Hackett Publishing Co., 2010).

On the question whether merely past or merely future times exist, see Thomas M. Crisp's "Presentism" and my own "Four-Dimensionalism", both in Michael J. Loux and Dean W. Zimmerman (eds), *The Oxford Handbook of Metaphysics* (Oxford: Oxford University Press, 2003), pp. 211–245 and 246–280. Both handbook articles contain copious further references to the literature on presentism and four-dimensionalism. The Time Traveler's speech from H. G. Wells's *The Time Machine* is reprinted in *Arguing about Metaphysics*, along with excerpts from Alan Lightman's provocative *Einstein's Dreams*. For H. Minkowski's comments about the relationship between space and time (quoted earlier in the chapter),

see his "Space and Time", in H. A. Lorentz et al. (eds), *The Principle of Relativity* (New York: Dover, 1908), 73–91. For an important recent challenge to the idea that debates like this one can be settled by appeal to physics, see Bradley Monton, "Prolegomena to Any Future Physics-Based Metaphysics", *Oxford Studies in Philosophy of Religion* 3 (2011): 142–165.

On the passage of time, two classic articles are D. C. Williams, "The Myth of Passage", reprinted in *Arguing about Metaphysics* and J. M. E. McTaggart, "The Unreality of Time" *Mind* 17 (1908): 457–474. Paul Horwich's "The Metaphysics of Now", reprinted in *Arguing about Metaphysics* presents a reconstruction of McTaggart's argument. C. D. Broad's reply to McTaggart's argument is in the section titled "McTaggart on the Unreality of Time", in his *Examination of McTaggart's Philosophy*, vol. 2, part 1 (Cambridge: Cambridge University Press, 1938). For a more recent argument against my claim that McTaggart's argument is "decisive" on the assumption that four-dimensionalism is true, see Bradley Rettler, "McTaggart and Indexing the Copula", *Philosophical Studies* 158 (2010): 431–434. The article by L. A. Paul on temporal experience that I discussed at some length is her "Temporal Experience", *Journal of Philosophy* 107, no. 7 (2010): 333–359. (The quotation from Williams that I used is from pp. 335–336 of her article.)

6

TIME TRAVEL

Time travel is a staple of contemporary science fiction, and the philo-sophically puzzling features of time travel are precisely those features that make for the most interesting science fiction stories. Nothing in contemporary physics seems to rule out the possibility of time travel. Nevertheless, many have thought that metaphysical considerations do. This chapter explores three of the most important philosophical problems about time travel.

The first is related to our discussion about non-present times. Many have thought that presentism is inconsistent with the possibil-ity of time travel. Suppose they are right. If presentism is true, this is a reason to reject the possibility of time travel. Then again, if time travel is possible, this is a reason to reject presentism.

The second concerns the famous "grandfather paradox". It seems clear that nobody can go back in time and kill their own grandfather before their grandfather has had the opportunity to sire children. For if one did do this, one would never come into existence to make the journey. But this is very strange. What would prevent one from doing such a thing? There is no known law of nature that prohibits it. Contrary to what is sometimes presupposed in time travel fiction, there are no "logic police" running around trying to ensure that time

travelers do not bring about incoherent states of affairs. So it looks as if the possibility of time travel implies that there are constraints on the behavior of time travelers that admit of no natural explanation.

The third issue pertains to backward causation. If time travel is possible, then some causes come *after* their effects. For example: Suppose you travel back to the beginning of the 20th century and then you start to think about some events in your own childhood. Those events are among the causes of your mental states in the early 20th century, and yet they will not occur until many years later. Worse, if time travel is possible then so are causal loops—causal chains that loop around on themselves, so that some events in the chain lie in their own causal history. But many people think that backward causation and causal loops are impossible. If they are, then time travel is impossible as well.

PRESENTISM AND TIME TRAVEL

The idea that presentism is inconsistent with time travel is intuitively quite compelling. How can one travel to the past if, as presentism affirms, there is no past to which to travel? The argument implicit in this question has been formulated by Simon Keller and Michael Nelson as follows:

6.1   Travel requires a departure point and a destination.
6.2   If presentism is true, there are no destinations for time travel.
6.3   Presentism is true.
6.4   Therefore, time travel is impossible.

Keller and Nelson refer to this argument, appropriately enough, as the "Nowhere Argument".

Our first clue that something may be amiss with this argument should be the fact that even ordinary temporal passage (e.g., the passage of time from your starting this sentence to finishing it) is a kind of "time travel". Says the presentist: There are no future events or times, but there *will* be, and we will somehow manage to make our way from one to the next. If this is true, then the Nowhere Argument

is unsound. Still, a presentist might think that time travel to the *past* poses a special problem. The reason is that, even if 6.1 is false, the following premise seems quite plausible:

6.5   Travel to a destination *d* is possible only if *d will* exist when one arrives.

Moreover:

6.6   Presentism implies that no past time (no time with the property *being past*) *will* exist.

On the supposition that presentism is true, these premises imply that travel to the past is impossible.

Despite its plausibility, I think that there is reason to question 6.5. It makes sense to suppose that travel requires that one's destination *will* exist so long as one is also supposing that travel is a process that carries one forward in time. Bracketing that assumption and allowing that some journeys might go backward in time, it is hard to see why we should continue to think that travel requires that the destination *will* exist. Shouldn't we rather say that travel to a location simply requires that the location exist when we arrive? If we do say this, then the revised Nowhere Argument is defused.

So the Nowhere Argument looks unpromising, as Keller and Nelson themselves point out. Moreover, there is a positive argument to be had for the conclusion that time travel is consistent with presentism. The positive argument is this: We can tell stories that are best characterized as time travel stories but which say nothing that seems to imply the denial of presentism. (Keller and Nelson offer further motivations for thinking that presentists should believe in time travel, but this is the only one I will discuss here.)

Consider the following story. Clare has a simple goal: to get rich. Her plan is to make two journeys. One trip will take her to December 1979, about six months after she was born, to make a $10,000 investment in Eaton Vance stock, the top performing stock between 1979 and 2005, the year when she began work on her time machine. The other trip will take her to 2005 to cash in the stock and put the proceeds in an offshore bank account. (According

to the *Eaton Vance Corp. 2004 Annual Report*, her 1979 investment would be worth over $10,000,000 by 2005.) She has very good reason to believe that her journey will be—or has been—successful. In February 2005, while still in graduate school, she received a package which contained (a) all of the paperwork, ostensibly signed by herself only a week before, for the offshore account, (b) a complete explanation of her time travel journeys and their outcome, and (c) detailed plans for building the time machine in which the journeys would be made. She knew that it could be a hoax—indeed, suspected as much at first. But as she read through the detailed notes on the time travel journey, written in her own hand and bearing on every page the unmistakable traces of her very own dry and self-deprecating sense of humor, she was soon a true believer. She made a digital copy of the plans for the time machine and burned the original. She then quit her job and withdrew a not inconsequential sum of money from the offshore account to fund what would become an eight-year project of building her time machine according to the specs she had been given. She also began the challenging task of rounding up $10,000 in pre-1979 US currency.

As with most time travel fiction, this story narrates some events leading up to a time travel journey, then "follows" the central character through the journey and back to the present. But if such journeys *really happened*, the relevant events would simply unfold in a linear way from the earliest moment in the overall sequence to the latest. Thus, the story about Clare (narrated in the present tense) would unfold as follows: At some point in 1979, a woman named "Clare" appears (perhaps in a strange vehicle, or holding an odd looking device) out of nowhere, carrying a bag filled with $10,000 in currency minted in 1979 or earlier. She sets up an investment account in her own name, invests the money, then disappears. (Meanwhile, elsewhere, there is a little baby named Clare who shares all of her DNA with this mysterious woman.) In February 2005, an extremely similar-looking woman, also named Clare (and with all the same DNA and memories as the one who showed up in 1979) appears out of nowhere and cashes in the investment. She then writes a note to another woman, younger but otherwise very similar, also named Clare, and sharing all of her DNA. She remembers *being* this younger woman and thinks of her as her younger self. The younger woman

receives the note, and … etc. Eventually, at what is now our present time, a woman named Clare—looking just like the woman who showed up in 1979 and 2005, and sharing all of that woman's DNA, activates a machine that she believes is a time travel device that will take her back to 1979 …

Nothing in the story obviously contradicts presentism; nor, it seems, do we contradict presentism if we *add* to the story the claim that, in fact, the woman who shows up in 1979 is the *same person* as the baby mentioned in the story, the woman who writes the note in 2005, the woman who receives the note in 2005 and the woman who steps into the time machine at the present time. That is, we can assume the story to be a true time travel story without presupposing that, when 1979 was present, any object or event from the 21st century existed; and we can also assume it to be true without presupposing that events from the year 1979 *now* exist. This is not to say that time travel stories will always be free of problems. For example, our story implies that adult-Clare and baby-Clare were, in 1979, the *very same person* but, on the other hand, *in different places at the same time*. That seems weird. But problems like this are not peculiar to time travel stories told under the supposition that presentism is true. Rather, they have to be dealt with by *anyone* who believes in time travel.

Presentism all by itself, then, offers no clear reason for rejecting the possibility of time travel. But what about the famous "grandfather paradox"?

THE GRANDFATHER PARADOX

Imagine that, instead of attempting a journey to 1979 in order to implement a scheme to get rich, Clare sets out on a journey to 1935 (when her grandfather was a ten year old boy) on a mission to assassinate her grandfather before he had the opportunity to sire children. Can she do it? On the one hand, she clearly cannot. If Grandfather is assassinated before he sires children, Clare is never born; if she is never born, she never travels to assassinate him; but, according to our story, she *does* travel to assassinate him; thus we have a contradiction. On the other hand, if we build enough information into the story, it seems that she clearly *can* assassinate him. Imagine that she is

an elite Navy Seals operative, traveling back to 1935 with the very best equipment that money can buy. She finds her grandfather sound asleep, all alone in a farmhouse in the middle of nowhere, recovering from an accident that has left him almost entirely physically incapacitated. It seems that, if anyone can do anything at all, Clare can kill her grandfather. The "paradox", then, is the fact that both claims about what Clare can and cannot do seem equally compelling, yet it seems that only one of them can be true.

David Lewis solves the paradox by denying that only one of the claims can be true. According to Lewis, ability claims are context-sensitive. Relative to what we know about Clare's abilities, equipment, and other conditions, Clare *can* kill her grandfather. Relative to what we know about the total story—including the fact that grandfather survives to sire one of Clare's parents—Clare *cannot* kill her grandfather. Hold one set of facts firmly in your mind and you will agree that she can; hold the other set of facts firmly in your mind and you will agree that she cannot.

I think that Lewis's response is successful at least to this extent: it shows us that there is no genuine *paradox*, i.e., no pair of absolutely contradictory claims that we have equally good reason to affirm. For, as he explains it, the claim that she can kill her grandfather doesn't really contradict the claim that she can't. The one claim talks about her abilities relative to one set of facts; the other talks about her abilities relative to another set of facts. However, I do not think that Lewis's response does justice to the depth of the puzzle.

Consider another example from Lewis: I *can* read French; that is, I have the ability to learn the language. But don't ask me to help you get around the streets of Paris, for I *cannot* read French; that is, I have no present knowledge of the language. Here too we have context sensitivity in our use of the word "can". But here, and in similar examples where we aren't even tempted to say that we face a paradox, our puzzlement vanishes when we say more clearly what we mean by each of the apparently contradictory ability claims. I lack the necessary skills and knowledge to read French; but I have the right sorts of cognitive capacities, opportunities, and so on to acquire the necessary skills and knowledge. This is not even a little bit puzzling. When we try to do the same in the case of the Grandfather Paradox, however,

the puzzle does not disappear. Clare does have the necessary skills and so on to kill Grandfather, and there are no obstacles to her doing so. But it is impossible that she do so. *This fact is what is puzzling.*

The problem, fundamentally, is that we seem to be confronted with a brute (unexplained) physical impossibility, and one which seems, oddly, to involve just a single individual. Of course, in one sense there is an explanation: a scenario in which someone kills their own, actual grandfather before he sires children is absolutely metaphysically impossible. If Clare were to kill this boy, her actual grandfather, she would be in such a scenario; therefore, it is impossible that Clare kill this boy. The trouble is that absolutely none of the other indicators of impossibility are present in the scenario. It seems that anyone else in the same position as Clare could have killed her grandfather. What, then, is so special about Clare?

In reply, there seem to be two ways to go, short of trying to rest content with the absence of further explanation. One way is to try to dissolve the puzzle by arguing that, not only is Clare unable to kill Grandfather, but *no one* would have been able to do so. This solves the problem by showing that the impossibility in question is principled; it is not confined to just one individual. The other way is to try to show that the alleged impossibility is not genuine—that, in fact, the past can be changed and Grandfather can be killed, even by Clare. My own preference is to go the first route, arguing that, between the earliest arrival point and latest departure point of a time travel journey, nobody is free. Others prefer the second route. It would take us too far afield to explore these routes in detail, however; so, for purposes here, I'll rest content having identified the two routes and will move on to the next topic.

BACKWARD CAUSATION AND CAUSAL LOOPS

A causal loop is a series of events such that each event in the series lies in its own causal history: Each event in a loop helps to cause *itself*. For example, the story about Clare that I told in the section on presentism includes a causal loop. The time machine that Clare uses at the present time is built from plans that she first received from her older self in the midst of her second journey with that machine. Here, it is only *information* (i.e., information about how

to build a time machine) that exists in a causal loop—the paper on which the plans are written is, according to the story, burned in 2005 and acquired sometime "between" the time at which it is burned and the time when Clare writes down the plans to give to her younger self. But some of the most interesting time travel stories present us with *objects* whose very existence is loopy. The object has no beginning in time and no end, and each "stage" of the object's career lies in its own causal history.

For example, in Robert Silverberg's famous story, "Absolutely Inflexible", we find a time machine whose entire career is a causal loop. In the story, a time machine is given to Mahler, the story's protagonist, by a time traveler who has been taken prisoner and put in a space suit for purposes of contamination avoidance. The time traveler seems to recognize him; but Mahler cannot identify the time traveler in part because his visage is obscured by the helmet of the space suit. In short order, Mahler, having been given the prisoner's time machine, initiates a time travel journey during which he receives a severe beating and is later returned to a time shortly before the time when he received the time machine. Upon his return, he is taken prisoner, put into a space suit so as to prevent him from contaminating others with any germs he might be carrying, and dragged before the magistrate … himself … to whom he gives the time machine. The *machine itself* has no life outside of this sequence of events. It is given to Mahler and taken on a journey which ends precisely where it started, with its being given to Mahler. It has no origin and no end. It exists in a causal loop.

Causal loops are deeply puzzling; many think that they are impossible. Some think that they are impossible because backward causation in general is impossible. But (unrestricted) time travel to the past is possible only if both backward causation *and* causal loops are possible. If you can travel to the past, and if there are no restrictions on where you can go in the past, then you can travel to places and times that allow you to do things that lie in the causal history of your own time travel journey. In such a case, every event in the journey will lie in its own causal history, and so some effects will be temporally prior to their causes.

Are causal loops, or backward causation more generally, impossible? Here I shall consider just two arguments for the conclusion that

they are. The first is a direct attack on the possibility of backwards causation, and rules out causal loops simply because the latter are possible only if the former is. (The argument is due to D. H. Mellor. The example that follows is not quite Mellor's; it is an adaptation of Mellor's which incorporates elements from Ted Chiang's science fiction story, "What's Expected of Us".) The second is an argument due to Aaron Segal, and rules out causal loops alone on the grounds that their existence is conceptually impossible.

Suppose I have a device in my office that (I claim) sends a signal an hour into the past to another device in the fire pit in my backyard— one which produces a spark that will ignite whatever fuel I have stacked up on the grate in my fire pit. I demonstrate my device as follows: First, I load up the grate with wood and a bit of gasoline. We then watch for a few moments, and, lo and behold, the device ignites the fire. Let the ignition time be t. A short time later we take a walk to my office, wait until exactly one hour past t, and I press the button. Thus my demonstration. Will you be convinced?

According to Mellor, the button-pressing event can be a cause of the earlier fire only if there is a non-coincidental, non-artificial positive correlation between pushings of the button at one time and the igniter's emitting a spark one hour earlier. But, he says, it is *impossible* that there be any such correlation. Why, we might ask? Notably, Mellor does not demonstrate the impossibility. Rather, he simply invites us to imagine trying to test for such a correlation. As he imagines it, we would *always* be able to falsify the correlation. That is, we would always be able to act so as to make it the case that the correlation does not hold. Suppose we load up the grate with fuel and we watch as a fire starts at noon. Instead of walking to the office and pushing the button, we could set out to falsify the correlation and simply stay home from the office. Or, in an effort to be more thorough, we could go to the office and stand guard outside to ensure that no one pushes the button at 1 pm. Or we could rush to the office and push the button repeatedly before noon, and then destroy it before 1 pm. Or … it seems there are all manner of things we could do to try to falsify the correlation.

Not every case of backwards causation would be so easily tested; and Mellor has given no argument for the claim that unfalsified claims of backward causal correlations could not possibly be true.

Still, I take it that his general idea is this: In cases we would know how to investigate, falsification is easy—absurdly easy. This fact by itself gives us reason to believe that the necessary correlations are impossible.

But not so fast. Consider again the claim that, in the cases we would know how to investigate, *falsification is easy*. What evidence, really, do we have that this is true? None, it seems, apart from (a) the fact that we have never observed the right sort of correlation between future causes and past effects, and (b) the *assumption* that, in fact, there simply could not be any such correlation. If we set aside (b), it is hard to see why we should agree that, of necessity, it would be easy to falsify *any* alleged backward causal correlation. Indeed, it is not hard to imagine having real trouble falsifying some such correlation. Suppose that every time we make, and carry out, a definite plan to go push the button at 1 pm, a fire starts at noon. Every time we post guards outside the door, no fire starts. Occasionally we observe a fire to start in the absence of any plan on our part, and we find, invariably, that the button is pushed at 1 pm—perhaps by a maintenance worker, or inadvertently by us. We rig up a device to drop a weight onto the button at random intervals and to record each time the weight is dropped; we set video cameras on the fire pit, post sentries around the pit and on the office door, and then wait—only to discover, once we check our video and our office records, that multiple fires started throughout the day exactly one hour before the weight was dropped onto the button. And so on. Obviously the fact that we can imagine all of this does not establish that backward causation is possible; but it does serve to undermine key premises in Mellor's argument for the conclusion that it is impossible.

What about causal loops? There can be no doubt that causal loops—and especially object loops (objects whose entire history consists in a causal loop)—are highly improbable. Consider Mahler's time machine, for example. Suppose that when Mahler receives it from himself, it is in pristine condition—not even a scratch—and is composed of certain particles, the ps. Between the time Mahler receives it from himself and the time when he hands it over to himself, whatever happens to the time machine in terms of nicks, scratches, minor alterations in its composition, and so on will have to be undone. If it is scratched on the time travel journey, the scratch will have to be removed by the journey's end. If, as he touches the

machine, certain particles are sloughed off, *those very particles* will have to be put back *in precisely the locations they were in when Mahler received the machine* by the journey's end. Massive improbability indeed! But so far this is no reason to think that causal loops are impossible.

Still, there is the following conceptual problem. The causal relation is commonly taken to be **irreflexive** and transitive. To say that it is irreflexive is to say that nothing is a cause of itself: for any event *e*, *e* itself is not among the causes of *e*. To say that it is transitive is to say that if *e1* is a cause of *e2* and *e2* is a cause of *e3*, then *e1* is a cause of *e3*. But if both of these things are true of the causal relation, then causal loops are impossible. Again, a causal loop is a casual chain in which at least one of the events in the chain lies in its own causal history. So, for example, suppose we have a causal chain such that *e1* causes *e2*, which causes *e3*, which causes *e4,* which causes *e1*. By transitivity, it follows that *e1* is a cause of *e1*, which violates irreflexivity. So argues Aaron Segal.

How serious is this problem for those of us who would like to believe in the possibility of causal loops? Ultimately, I think that it is not serious at all. Self-causation is surely rare, surely improbable; but I see no reason to think that it is ruled out by the very concept of a cause. Although the world of fiction is filled with impossible stories, stories that straightforwardly violate *conceptual* truths are rare, and they would, in any case, strike us immediately as incoherent. (Just imagine trying to get your mind around a romance story in which a woman has an affair with a married bachelor.) But time travel stories that incorporate causal loops are ubiquitous; and, though some are manifestly incoherent, not nearly all of them strike us this way. This is very surprising if our concept of causation rules out causal loops in the way that our concept of bachelorhood rules out the possibility of married bachelors. Although this is not proof that the objectors are incorrect, it at least provides grounds for insisting that they argue, rather than merely assert, that "our" concept of causation includes transitivity and irreflexivity. So far as I am aware, no convincing argument for this conclusion has yet been given.

FURTHER READING

David Lewis's "The Paradoxes of Time Travel", *American Philosophical Quarterly* 13 (1976): 145–152 is an absolute must-read for anyone interested in the metaphysics

of time travel. *Arguing about Metaphysics* reprints a couple of excellent short stories that illustrate the idea of a causal loop: Robert Heinlein's "—All You Zombies—" and Robert Silverberg's "Absolutely Inflexible". I also mentioned Ted Chiang's story, "What's Expected of Us", which is reprinted there too. For a good book-length introduction see Ryan Wasserman's *Paradoxes of Time Travel* (Oxford: Oxford University Press, 2018).

On the topic of presentism and time travel, I referred to S. Keller and M. Nelson, "Presentists Should Believe in Time-Travel", *Australasian Journal of Philosophy* 79, no. 3 (2001): 333–345. See also Theodore Sider, "Traveling in A and B Time", *The Monist* 88 (2005): 329–335. D. H. Mellor's argument against causal loops is to be found in his *Real Time II* (New York: Routledge, 1998). Also on causal loops, see Aaron Segal, "Half-Hearted Humeanism", *Oxford Studies in Metaphysics* 9 (2014), and Richard Hanley, "No End in Sight: Causal Loops in Philosophy, Physics and Fiction", *Synthese* 141, no. 1 (2004): 123–152. Hanley defends at length the idea that causal loops are merely improbable, not impossible.

On changing the past, see Peter van Inwagen, "Changing the Past", *Oxford Studies in Metaphysics* 5 (2010): 3–28, and the reply by Hud Hudson and Ryan Wasserman, "Van Inwagen on Time Travel and Changing the Past", *Oxford Studies in Metaphysics* 5 (2010): 41–49. My own view that nobody is free between the earliest arrival point and the latest departure point of a time travel journey is defended in my "Time Travelers Are Not Free", *Journal of Philosophy* 112 (2015): 266–279.

7

SUBSTANCE

What is the difference between a vivid dream and the world outside? For mere mortals, even the most vivid dream is still a pale imitation of reality. But we can imagine that difference being removed. What further difference would there be? Our dreams are typically short, and display precious little continuity with one another, especially by comparison with our waking lives. Few of our dreams are governed by natural laws in the way that the real world is. One could hardly do good science even within the longest of dreams. Indeed, one can hardly make *any* plans or predictions in one dream that will carry over reliably to the next dream, or even to the next moment. Our agency is diminished in our dreams, too. If we deliberate or act with intention, here again we find only a pale imitation of what transpires in waking life. But we can imagine all of these differences removed as well. We can imagine, that is, a genuine dream that is no less vivid than waking life, just as law-governed, extraordinarily long, and something within which we can plan and predict in just the ways in which we do in waking life.

So imagine it.

What differences remain now between dream and reality?

The difference is substance.

In ordinary life, when we talk about substances we are usually talking not about objects but about undifferentiated masses of matter. Ask your friend to identify the substance on your desk and they will come looking for something more like a smear of peanut butter, or a powder of unknown origin; they won't look for any particular kind of object. As we saw in Chapter 3, however, philosophers have something different in mind when they talk about substance. What they have in mind is an object that has at least the following two features: *independence* and *unity*. Dream-objects (and dreams themselves) lack the first feature. They depend for their existence and character entirely upon our minds. Arbitrary combinations of objects lack the second feature: they have no real unity, which is why many people do not think of such things as genuine composites.

Needless to say, however, not everyone believes in substances as I have just characterized them. More exactly, not everyone believes that there is a distinction to be made between *genuine* ways of unifying objects and merely apparent ways; and not everyone believes that there is a privileged class of independent things that is somehow set apart from all other things, the dependent objects. Those who do believe in a category of substance, furthermore, differ widely as to what things count as substances and what their specific character is like.

In the first two sections of this chapter, we shall examine three of the most important theories about substance: *bundle theory*, according to which nothing satisfies the traditional definition of substance; *substratum* theory, according to which the substances are bare particulars that somehow underlie all of a thing's properties; and *hylomorphism*, according to which a wide variety of familiar material objects count as substances. Finally, in the third section, we turn our attention toward God, who is identified in the Aristotelian hylomorphic tradition as a substance *par excellence*, the ground of all being.

BUNDLES VERSUS SUBSTRATA

Objects have attributes; nothing could be more mundane. Suppose we are realists about attributes: We think that attributes themselves

are abstract objects of a certain kind—perhaps universals, perhaps tropes. One might now ask what *having* an attribute will involve.

Before addressing this question, let us first say something about the term "attribute". Attributes are just properties. Sometimes philosophers use the term "attribute" to signal their adherence to a particular conception of properties—for example, that properties modify the things that have them rather than belonging to them as constituents (as they would if properties are tropes) or having them as members (as they would if properties are sets). Sometimes, too, it seems that philosophers talk about attributes rather than properties because they are trying to focus on intrinsic properties, or just those properties that would exist if a sparse theory of properties is true. As mentioned earlier, in this book the term "attribute" is being used as if it is synonymous with the term "property". But for the remainder of this chapter I shall presuppose realism rather than nominalism about attributes. More specifically, I shall presuppose that attributes are either universals or tropes.

Given this assumption, there are, broadly speaking, two main answers that one might give to the question of what it means to have an attribute. One might say that attributes are constituents of the objects that have them; or one might say that attributes are in some sense external to the objects that have them. (Reminder: I have been using the term "constituent" throughout this book as if it is synonymous with the term "part".)

The difference between these two answers marks the difference between what are sometimes called *constituent ontologies* and *relational ontologies*. The labels are not entirely apt. For one thing, constituency itself seems like a relation; so there is really nothing any less "relational" about constituent ontologies than about their rivals. But let us set this and other worries aside for now. Despite the labels, the main difference between constituent and relational ontologies does not so much have to do with one involving relations and the other not. Rather, it has to do with the family of metaphors that proponents of each kind of ontology tend to use to talk about property-having (whatever exactly it amounts to).

People who accept a constituent ontology tend to say that properties are somehow *in* their bearers, or that they *inhere* in the things that have them. They say that properties are *constituents* of the things that

have them. Proponents of a relational ontology tend to talk about objects *participating in* properties, or they talk about properties being *transcendent* rather than *immanent*, or *external* to their bearers rather than *in* them as constituents.

Those who endorse relational ontologies typically maintain that attributes are just as independent as—maybe even more independent than—objects in the material world. On that sort of theory, attributes *are* substances. Although they are attributes, they also *have* attributes (Plato, for example, thought that properties at least exemplify themselves). They are also basic, or independent things on which other things depend. Having no parts, they are also unified in the deepest sense of the term. Within Plato's relational ontology, properties are the only substances. But other relational ontologies might either hold them to be on a par with material substances or reject the substance–attribute distinction altogether.

Turning to constituent ontologies, we find two further options. First, to borrow a familiar classroom metaphor, we might think that attributes relate to substances as pins relate to a pin-cushion: like pins, the attributes are somehow "stuck" to the substance, and the substance (i.e., the pin-cushion) is, in and of itself, a thing with no attributes. On this view, then, a substance is the sort of thing from which—conceptually speaking, anyway—all attributes could, in principle, be removed. For this reason, this first view is commonly referred to as the *bare particular*, or *bare substratum*, theory of substance. One might endorse a bare particular theory and treat substrata and their properties as on a par in terms of their independence and unity. More commonly, bare particulars are seen as enjoying a kind of independence that attributes do not. Attributes are instead seen to depend for their existence on their attachment to bare particulars, but not vice versa. Thus, on this sort of view, there are no unexemplified attributes, and bare particulars are the only substances.

Second, we might think that the relation between objects and their attributes is like the relation between a bundle of sticks and the sticks themselves: i.e., there is nothing more to the thing than the sticks. Notice that one might think of a bundle of sticks as an object that has the sticks as parts; or one might think of the bundle as a **mere plurality**—just some sticks which do not compose any further thing. Taking the first way of thinking, bundle theory is the

view that everything is either an attribute or a bundle of attributes. Taking the second way of thinking, bundle theory is the view that, strictly speaking, everything is an attribute (since mere pluralities are not objects), and that what we *take* to be familiar objects are really nothing but mere pluralities of attributes. Either way, this is not an ontology according to which there are independent, unified things which *have* attributes. Bundles have attributes, but, being built up out of their attributes, they depend on their attributes rather than the other way around. Attributes, on this view, are independent; but they do not have attributes, for otherwise they too would be mere bundles. So it is not really a theory about the relation between substance and attribute. It is, rather, a denial of substance altogether.

Both the bare particular theory and the bundle theory have their fair share of problems. Consider first the bare particular theory. The first and most obvious objection concerns the characterization of bare particulars themselves. It seems to be central to the view that bare particulars are *bare* (i.e., they have no attributes) and *particular.* But now we face a dilemma. Either we must say that bare particulars have the *attributes* of being bare and particular in such a way that, unlike every other attribute stuck to the cushion, those two could not possibly be stripped away; or we must say that the predicates "is bare" and "is particular" apply to and characterize bare particulars despite the fact that there are no attributes corresponding to the predicates. The first horn of the dilemma is problematic because the very concept of a bare particular is the concept of something from which *all attributes* can, in principle, be stripped away. Thus, it looks as if the very idea of an attribute like *being bare* is incoherent: anything that has the attribute could not possibly have the attribute. But the second horn of the dilemma is also problematic. Bare particular theory is, by its very nature, committed to realism about attributes. But if we really thought that it was possible for predicates like *being bare* and *being particular* to characterize objects in the absence of corresponding attributes, the motivation for believing in attributes would seem to be lost.

Another problem is that bare particular theory seems to result in an unreasonable multiplication of subjects of predication. Suppose that Socrates is seated and is thinking about his wife, Xanthippe. According to the bare particular theorist, this fact obtains because

Socrates's bare particular has at least two attributes stuck to it: *being seated* and *thinking about Xanthippe*. But now we face another dilemma. Either Socrates is identical to his bare particular or not. Suppose Socrates *is* identical to his bare particular. Socrates himself is not bare; he has lots of attributes and some of those attributes, like being capable of thought, are such that, were they stripped away from him, Socrates would cease to exist. So that option is not viable. Suppose then that Socrates *is not* identical to his bare particular. Now we have a different sort of problem. Socrates, as we have said, is seated and is thinking of Xanthippe. But it seems that his bare particular is as well; for, after all, those attributes are stuck to the bare particular— the bare particular *has* them. But if that is right, then there are two things seated in the same place at the same time, both thinking of Xanthippe and both, indeed, thinking that Xanthippe is their wife!

One might try to respond to this problem by arguing that Socrates has the attributes just mentioned in one way whereas the bare particular has these attributes in a rather different way. If we say this, we will identify one way of having attributes as primary, and another way as derivative. There is some precedent for this. For example, an apple is red. Strictly speaking, just the surface is red; but we talk as if the whole thing is red. So we might say that the surface has redness in the primary way whereas the whole apple has redness derivatively— it counts as red simply *because* the surface is red. Likewise, one might think, with Socrates and his bare particular: One has *thinking of Xanthippe* in the primary way; the other has that attribute only derivatively. Whatever its merits, however, this solution does not do much to mitigate our problem. For, on this view, Socrates and his bare particular still *have* the attributes in question. There are still two things seated in Socrates's chair, and still two things thinking of Xanthippe and thinking that she is their wife. That one of the things is only derivatively thinking of Xanthippe does not seem to make this consequence any more palatable.

Bare particular theory has problems. Unfortunately, so too does the bundle theory. First, we might ask what exactly *bundling* amounts to. Even if bundles are mere pluralities, not every plurality of attributes comprises a bundle. (Half of my attributes and half of yours, for example, do not together comprise a bundle.) Suppose that attributes are universals. They are, in other words, necessarily existing abstract

objects that can be instantiated in multiple places at once. What could it possibly mean to "bundle" some of those sorts of things together?

We might say that *n* universals are bundled together when they are instantiated in exactly the same place at the same time. But that won't quite work. Consider, for example, a bowl of spaghetti. According to the bundle theory, the bowl of spaghetti is a bundle of universals. But the colors and flavors of the bowl of spaghetti and the various textures are not all instantiated in exactly the same places. The tomato flavor of the sauce isn't instantiated in the same place as the hardness of the bowl, for example; and the mass of the entire composite isn't instantiated in exactly the same region as the color of the pasta (or even the same region as the composite color of the bowl of spaghetti taken a whole). And, of course, similar problems arise if we take attributes to be tropes rather than universals. The color tropes for the bowl of spaghetti wouldn't be in exactly the same place as the flavor tropes, or the hardness trope, or the mass trope. The typical response to this problem is to introduce a new relation—for example, "compresence"—and to say that bundling occurs whenever tropes or universals stand in that relation to one another. But, of course, that doesn't explain anything; giving that sort of answer really does little more than ask us to rest content with the mystery of bundling.

Second, we might ask how bundling, whatever exactly it is, manages to generate *concrete* objects. Suppose that attributes can be located in spacetime. It doesn't really matter now whether we think of attributes as tropes or universals; either way, they are going to be abstract objects, which is all that we need to generate the problem. Now you might wonder why we would suppose that attributes can be located in spacetime. After all, they aren't concrete objects; and it certainly seems odd to say, for example, that the tomato flavor of your spaghetti is sitting right there on the table in front of you. (Then again, if it isn't there on the table, where else might it be? And if the flavor were literally nowhere, then how could you ever get it into your mouth?) But the reason we should feel comfortable here assuming that attributes can be located is that bundle theory seems to require it. For, again, whatever exactly bundling attributes amounts to, it seems obviously to require that attributes can be put together in some locations without being put together in others. So suppose we have an attribute A1 located in a particular region, R. If just one

attribute is there, we obviously do not have a concrete object; we just have the attribute, which is abstract. So add an attribute, A2. Does that give us a concrete object in R? I do not see how it could. Where would the *concreteness* come from? So add a third attribute, A3. Does that now give us a concrete object in R? Again, it is hard to see how it could. And so on.

One might say that concreteness is itself an attribute; so the concreteness comes only when we add *that* attribute to the bundle. But this will solve the problem only if concreteness is itself concrete. Initially, one would not think that it is—concreteness is an attribute like any other, and attributes are supposed to be abstract on this view. If we do say that concreteness is concrete, there is some pressure to say that *all* attributes exemplify themselves. The reason is that it would be hard to find any principled reason for saying that some attributes exemplify themselves but others do not. If we do say that all attributes exemplify themselves, however, then we confront further oddities: humanity is human (thus, we have an abstract human being); redness is red (so an abstract object has a color); beauty is beautiful (so an abstract object with *no other features* somehow counts as beautiful); and so on. None of these claims is obviously incoherent; but neither do they seem to be the sorts of commitments one should be happy to take on. But if we do not say that concreteness is concrete, then it is very hard to see how we could ever get a concrete object out of a process of "stacking" abstract objects together in a region of spacetime.

Third, if the bundle theory is true, it is hard to see how objects could ever have different properties than they do at any given time. To take an analogy: Suppose you have a bouquet of roses. Now suppose I replace a rose with a carnation. Do you have the same bouquet? Obviously not. Suppose instead I replace one rose with another rose. Now do you have the same bouquet? Again, it seems not. A bouquet of roses, which is just a bundle of flowers, seems to have its constituent roses *essentially*. That is to say, it is a necessary truth that that particular bouquet exists only if those particular roses are parts of it. Likewise with bundles of attributes. Replace one attribute with another and you have a different bundle. But if an object is nothing more than a bundle of attributes, then adding or removing an attribute to or from the bundle will leave you with

a different object. For example: Suppose that Socrates is seated. The bundle theory implies that Socrates is identical to a bundle (call it "B1") that has among its constituents the attribute *being a seated man*. If Socrates were to stand up, however, Socrates would be identical to a bundle (call it "B2") that has among its constituents the attribute *being a standing man*. B1 is not identical to B2; thus, Socrates cannot be identical to both of them. Thus, either Socrates cannot stand up or the bundle theory is false.

Unsurprisingly, plenty of philosophers have tried to defend bundle theory from these and other objections. Each of these philosophers, Laurie Paul most notably, moves the bundle theory in the direction of saying that, ultimately, *everything is a property*. I think that it would be premature to say that this striking thesis, too, is a consequence of bundle theory; for perhaps there are yet ways of defending the theory that do not move in that direction. But for those who would rather not find themselves believing that everything is a property, bundle theory should be approached only with caution.

HYLOMORPHISM

The bundle theory and the bare particular theory might be seen as agreeing on one basic principle. *All attributes are alike* in the following way: either they all have to "stick" in the same way to an underlying individual subject, or none of them have to stick to an underlying subject at all. The bundle theorist takes the second option, denying substance altogether. The bare particular theorist takes the first, positing intrinsically bare individuals as the underlying subjects. The bare particular theory foundered, as we saw, in part because it is hard to see how anything can be intrinsically bare. But perhaps that problem is just an outgrowth of the idea that all attributes have to stick *in the same way* to an underlying *individual* subject. The theory of substance known as "hylomorphism" is not committed to this basic assumption (or, at any rate, not *every* proponent of hylomorphism commits to it), and so perhaps it offers a way forward.

The term "hylomorphism" derives from two Greek roots: *hyle* which means "matter", and *morphe*, which means "form". The core idea is that every material substance is a structured entity with two constituents, matter and form. Form is a special kind of attribute; it

relates to its underlying subject in a way somewhat different from the way in which other attributes relate to their underlying subjects. Matter, likewise, is a special kind of subject; it is neither a bare particular nor a substance, which are the only other kinds of subject with which we are thus far acquainted.

The hylomorphic theory of substance has its origins in Aristotle. But it was also the dominant theory of substance in the Middle Ages, and underwent a great deal of modification and development in the work of the great medieval philosopher-theologians, most notably St Thomas Aquinas. A variety of early modern philosophers—Descartes and Leibniz (1646–1716), for example—also had their fair share to say about matter, form, and substance; and what *they* had to say about those concepts was influenced in part by developments in the natural sciences. Skipping ahead to the 20th and 21st centuries, hylomorphism is undergoing something of a revival in contemporary metaphysics. Among the most important figures in that revival are Jeffrey Brower, Kit Fine, and Kathrin Koslicki; but their own versions of hylomorphism are informed by still other influences, including the formal logics developed in the early and middle parts of the 20th century, and various aspects of contemporary scientific theory. It is thus hard to say much that is uncontroversial about the *general* nature of hylomorphism; each proponent seems to have their own particular take on the theory. In fact, what we have at this point is not so much a single theory as a philosophical tradition that has given rise to a family of related theories.

As I see it, the central elements of hylomorphism are the following three theses:

(H1)   Objects are compounds of matter and form.
(H2)   An object's form is its *nature* or *essence*.
(H3)   Natures are principles of unity and development.

I will discuss each thesis in turn.

As I said earlier, matter is a special kind of subject: neither a bare particular nor a substance. What, then, is it? Perhaps the simplest characterization is that matter is whatever it is that a substance is made of. Consider, for example, a clay pot, created on a potter's wheel. Few in the Aristotelian tradition would actually regard clay pots as

substances. As we shall see shortly, they don't have the right kind of "form". But never mind that for now; even Aristotle used examples like this to illuminate his concepts. When a potter forms a clay pot, the pot is made from an amorphous lump of clay. Prior to being shaped into a pot, the lump exists and is *potentially* a pot. The activity of the potter takes this pre-existing stuff and makes it *actually* a pot by imposing a certain shape upon it. Thus, there are at least two things that we can say about the clay. First, the clay is the underlying subject that receives the form *being a pot*, thereby bringing the pot into existence. Second, the clay is that aspect of the pot that, apart from the form, is *merely potentially* a pot. For this reason, in Aristotle's metaphysics matter is associated with potentiality. It possesses the potentiality that is actualized by form, and it is sometimes said to be something that exists in potentiality. Form, by contrast, is associated with actuality. It actualizes the potentiality inherent in matter, and it exists in actuality.

We can now give a more general and accurate account of matter. Let "K" be a term that refers to a substance kind, like *human being* or *cat*. (In the Aristotelian tradition, living organisms are paradigm substances.) So, for example, saying that *x* is a K will be grammatically just like saying that *x* is a human being, or that *x* is a cat, or whatever; and talk about K-ness will be grammatically like talk about *humanity* or *felinity* or whatever. Given this terminology, we can characterize matter as follows: The matter for a K is whatever (a) is the underlying subject of the form K-ness and (b) is, apart from its K-ness, merely potentially a K. There is room for disagreement about the "level" at which we find the matter of a thing. In a human being, for example, we might think of the matter as the mass of tissue and bone and other stuff that makes up the human being; or we might think of it instead as a mass of "lower level" items—cells, perhaps, or particles—that are arranged human-wise. Similarly for trees, horses, and so on.

Suppose we say that the matter for a human being is the mass of tissue, bone, etc. that makes up the human being. Now we face a further question. Is this mass itself a compound of matter and form? Here again there is room for disagreement. There are three main options. One option is to say that (e.g.) the matter for a human being is itself an individual substance whose form is something like *being a mass of tissue, bone, etc.* Few in the tradition would take this option.

A second option is to say that what we are calling "the mass of tissue, bone, etc." is a *mere plurality* of smaller substances (cells, fundamental particles, etc.)—not itself an individual thing at all. A third option is to say that the matter for a human being is an individual pseudo-substance. According to this third option, there is a hierarchy of material objects: some are substances in the primary sense of the term; others are substances only in a secondary or perhaps analogical sense of the term (which is why I call them "pseudo-substances"). Living organisms, for example, would likely count as primary substances; artifacts and mere lumps and so on would count as substances in the secondary or extended sense of the term. The primary substances, then, would have matter and form in the primary senses of those terms; the other objects would have matter and form in secondary or extended senses of those terms. Finally, a fourth option is to say that, in fact, the matter for a human being or tree or whatever is not a mass of tissue, bone, and so on but instead a mass or portion of what has typically been called *prime matter*—stuff that, in and of itself, has no qualities whatsoever. (Those who believe in prime matter usually say that some, but not all, objects have prime matter as their matter. They then take one of the other three options in accounting for the matter of objects that are not compounds of prime matter and form.)

The idea of prime matter is about as puzzling as the idea of a bare particular; but it is better off in at least one respect. A bare particular is supposed to be a substance—a genuinely unified and individual thing, a basic entity on which other things depend for their existence. As a substance, the bare particular seems to be the sort of thing which ought to have an intrinsic nature. That is, we ought to be able to say *what it is* to be a bare particular. And there is precisely our problem: To be a bare particular is, if anything, to be an individual thing that has no intrinsic nature. The idea seems contradictory. Prime matter, on the other hand, is not supposed to be the sort of thing that has an intrinsic nature. Indeed, in and of itself "it" is not a *particular* thing, nor is it any *sort* of thing at all. It has no independent existence whatsoever, and it has no intrinsic features. This is a hard doctrine, but it is not manifestly incoherent in the way that the idea of an intrinsically bare particular seems to be.

Still, prime matter is baffling. The problem with trying to explain the idea is that it seems that we can only do so under the assumption

that prime matter is some particular kind of thing that has certain kinds of features. That, after all, is how we describe *anything*. So if we say that prime matter isn't a particular, it belongs to no kind (not even the kind "prime matter"!), and has no intrinsic features, we are stuck—we have no idea what we are talking about. For this reason, quite a lot of people have been reluctant to believe in prime matter.

Enough about matter for now; what about form? We have already had a glimpse of what forms are supposed to be. They are more than mere shapes; they are complex organizational properties. Furthermore, as I have already said, forms are associated with actuality—among other things, they actualize the potentiality in matter. Accordingly, forms are typically also said to be the things that make something *what it most fundamentally is*. They are principles of unity, and they are the essences, or natures, of the things that have them. The nature or essence of a thing, in the Aristotelian tradition, is very roughly, a complex property which explains a large number of scientifically interesting features of the objects that have it—i.e., features of their outward appearance, behavior, natural development, and so on—and provides a basis for classifying those objects together as members of a kind. Aristotelian kinds are definable in terms of necessary and sufficient conditions. In other words, there are criteria for membership such that satisfying all of those criteria suffices to make something a member of the kind, and nothing can count as a member of the kind unless it satisfies all of those criteria. Furthermore, Aristotelian natures are also principles of change and development. That is, it is by virtue of having the particular nature that it has that a thing changes and develops in the ways that are characteristic of things of its kind.

Return again to the example of a human being. Apart from their form, there is nothing really to a human being besides some tissue, bone, and other materials. Those materials on their own, obviously, do not make a human being; they have to be *arranged* in the way that is characteristic of human beings. And by "arranged" I do not just mean that they have to be laid out in the shape of a human being; for arranging tissue, bone, and other materials in a human *shape* does not suffice for having a living, functioning human being. The arrangement, thus, has to include complex causal connections between the bit of tissue, bone, and so on. This is why I said earlier that form isn't mere shape.

Once the form is imposed on the tissue, bone, and so on, we have *a single human being*; prior to that, we don't have much by way of genuine unity at all. There is, for example, no system relative to which the bits of tissue, bone, and other materials have functions (or, at any rate, there does not seem to be). Those things are not playing any particular functional role in a larger whole. So this is one indication that, absent the form, there is no real unity to the matter. It is in this sense, then, that the form is a principle of unity: It takes us from having a mere plurality of substances, or a pseudo-substance, or just some stuff that is not yet a particular thing, to having a single unified substance; and *functional arrangement* is at least an indicator of, if not a necessary condition for, the presence of genuine unity.

Note that it is plausibly the functional arrangement of the parts of a substance that determine how the substance is supposed to change and develop over time. A human zygote has its parts functionally arranged in a particular way. Again, it is not just that they have a certain shape; rather, they have been put in certain causal relations to one another. Those causal relations, furthermore, together determine how the zygote will develop into an adult human being. It is in this way that natures play the role of principles of change and development.

I said earlier that forms are special kinds of attributes. We can now be more specific about the ways in which they are special. First, forms join with matter to generate substances. Furthermore, they *inhere in* matter without *characterizing it*. Humanity, for example, inheres in Socrates's matter. But it doesn't characterize Socrates's matter; his matter isn't human. Rather, humanity characterizes Socrates. Other attributes, on the other hand, do not join with matter to generate substances; and they characterize that in which they inhere. If Socrates is seated, seatedness inheres in him. It also characterizes him. And the inherence of seatedness in Socrates does not generate a substance. Second, forms are kind properties: they are properties the having of which determines what kind of thing an object is. Insofar as they determine an object's kind, they also determine its persistence conditions.

There is one final point about form that deserves our attention. Throughout this discussion I have been presupposing that all forms are substantial forms—i.e., all forms are the forms of substances; all forms correspond to substance-kinds, like *humanity* or *felinity*. However,

some Aristotelians think that just about any attribute can *play the role* of form. Thus, for example, some will say that, when Socrates sits down, a new compound, Seated Socrates, comes into existence. In Seated Socrates, *Socrates* plays the role of matter and the attribute *seatedness* plays the role of form, generating a new object (though not a new substance), Seated Socrates, that exists for just as long as Socrates is seated. Aristotle himself used examples like this. Some people recoil at the thought of objects like Seated Socrates. But I think that in fact most of us believe in such things. You probably believe that *Socrates's being seated* names an event. Why not say instead that it names an object, Seated Socrates? You probably believe you can make a fist. Your fist is perhaps not a substance. But it seems to be a matter-form compound whose matter is your hand and whose form is a particular fist-like shape. In any case, what this means is that what I have said about the way in which forms are special applies *only* to substantial forms, and not to whatever other attributes might happen to be playing the role of form. To avoid confusion, however, I will continue in this book to use the unqualified term "form" to refer only to *substantial forms*. When I want to talk about the constituents of things like fists, or Seated Socrates, I will be careful to speak either of "non-substantial forms" or of attributes *playing the role* of form.

We have now finished discussing the theories of substance on which I wanted to focus in this chapter. I want to end the chapter by briefly considering a topic that naturally arises in connection with discussions of substance—the existence of God. The topic arises because substances are supposed to be basic entities, which naturally leads into the question whether there is a single *most basic* entity (God?) on which everything else depends, or a plurality of such entities, or an infinite chain of dependent things.

GOD

For much of our discussion in this chapter we have focused on the idea of substances as things whose parts are genuinely unified. Now it is time to focus more on the idea of substances as things on which other things depend. In the Aristotelian tradition, as we have seen, living organisms and perhaps other composites are counted as

substances. Forms join with matter to produce hylomorphic compounds which are, then, the subjects in which all other attributes inhere. Furthermore, it is typical within the Aristotelian tradition to say that attributes depend on substances for their very existence. There are no uninstantiated attributes; indeed, attributes do not even *exist* in the primary sense of the term.

One might notice, however, a couple of problems with this framework. First of all, doesn't it seem that material substances are dependent things after all? True, their attributes might depend on them in the sense that, had there been no such things, then the attributes distinctive of material substances would not have existed. But, as hylomorphic compounds, these things seem, in turn, to depend on their constituents—on their matter and form. Second, material substances are the sorts of things that *come into existence* and *pass out of existence*. But if *they* are the things on which everything else depends, that would seem to imply that (a) everything is destructible, and (b) the universe contains an infinitely long sequence of generated and destructible objects.

These claims together, however, seem puzzling. For if everything is destructible, one would not expect there to be an infinite sequence of anything; one would rather expect that, after a suitably long time, any sequence of objects would end with the ultimate destruction of everything. It would be sheer coincidence, it seems, if a sequence of destructible beings lasted forever, and it seems we should not believe that such remarkable coincidences occur. So for this reason it seems that we shouldn't think that the universe contains an infinitely long sequence of destructible things. A sequence of destructible things that extends infinitely backward in time would surely have come to an end long before now; so our very existence seems to be evidence that there is no such sequence.

It seems, then, that the sequence of material substances must be finite; it must have a beginning and (short of some miracle) an end. Let us ask, now, about the beginning. Is that the culmination of some other sequence of changes extending infinitely backward in time? Or is it the result of the activity of some cause that is not a material substance? Or did the finite sequence of material substances pop into existence by chance?

Let us rule out the chance hypothesis. I do not know how to show that hypothesis to be absurd; but it is hard to imagine how *anything* could simply pop into existence by chance.

Suppose we say that the *beginning* of the sequence of material substances is the culminating result of some other sequence of changes extending infinitely backward in time. On this picture, there is a point in time when material substance comes into existence; but prior to that time and, indeed, extending infinitely backward in time, there is still change of some sort and, indeed, one or more of these changes ultimately resulted in the generation of material substance. For convenience, let us refer to the changes that are taking place before the generation of material substance as the "primordial changes". We may now ask what are the subjects of these primordial changes. Change occurs only if something changes; so what are the things that are undergoing the primordial changes? They cannot be material substances; for the primordial changes predate the existence of material substance. But what else is there? That is our first question. Our second question is whether each primordial change has a cause, so that there is an infinite sequence of causes, or whether there is some first cause.

Answering the first question provides a response to an earlier question that we have thus far left hanging. Toward the beginning of this section, I asked whether it seems that material substances themselves depend upon their constituents. For rather complicated reasons that are best ignored in this discussion, those writing in the hylomorphic tradition have been reluctant to elevate *matter* to the level of substance. But form—which, as we have seen, is traditionally associated with *actuality*, which constitutes the very *nature* of a substance, and which is also traditionally regarded as a principle of change and development—does have some claim to the status of substance. One can be human apart from any particular "portion" of matter; but one cannot possibly be human apart from humanity itself. Furthermore, as a principle of development, the form plays an important causal role in bringing a substance to and maintaining it in its mature and properly functioning state. Thus, it seems that material substances depend on their form in a way in which they don't depend on their matter. Furthermore, forms have a kind of independence that prime matter, at any rate, does not. Like familiar substances, and unlike prime matter, forms admit of intrinsic characterization; they are things in their

own right (universals or tropes). Finally, there is no reason to doubt that forms have their own internal unity. Thus, there seems to be good reason to accord them the status of substance. If we do, we have an answer to the question of what *besides* material substances there might be. There might be something like a pure form.

Note, however, that I say only "something like" a pure form. If forms are truly attributes, and *if* attributes truly depend on their instances, then there cannot be an unexemplified form. Thus, if immaterial substances pre-exist material substances, they might well be form-like. Following Aristotle, for example, we might characterize such a thing as *pure actuality*, and regard it as a primordial cause, a principle of change and development that explains other changes in the universe. But we could not say they are pure attributes (or anything like attributes) without giving up the principle that forms, like other attributes, depend upon their instances.

As to the second question, the answer depends upon whether one thinks that it is sensible to believe in an infinite sequence of causes. Aristotle did not, and many in his wake—theists and atheists alike—have not. Contemporary atheists who follow Aristotle on this point typically say that the "first cause" is just the initial singularity from which emerged the Big Bang. This is a hard doctrine, however, within an Aristotelian substance ontology. For if the initial singularity is a material substance, then it will be the sort of thing that is generated and destructible, which then means that we can sensibly ask where it came from. If it is not a material substance, then whatever it is, it is neither a genuine unity nor an independent thing. It is thus the sort of thing whose existence demands an explanation.

Aristotle, as I have said, thought that there was a first cause; and, on his view, the first cause *had* to be not a material substance but rather *pure actuality*. This, as we have just noted, would make the first cause something like a form. Aristotle went on to argue that the first cause had to be eternal, immovable, and, indeed, good (since goodness is a principle of motion, and the first cause is the ultimate explanation of all motion). Ultimately he identified the first cause with God.

The argument I have just presented is a fleshed out reconstruction of a line of argument that can be found in Book 12 of Aristotle's *Metaphysics*. The basic ideas in this argument have, as many readers will know, been developed in various different ways (and typically

with greater reliance on claims about causation, explanation, and dependence) throughout the history of philosophy—most notably in the Middle Ages by the Islamic philosophers Al-Kindi (801–873) and Al-Ghazali (1058–1111), and by St Thomas Aquinas. The versions developed in the Middle Ages have come to be known as versions of "the Cosmological Argument".

Although the conclusion of the Cosmological Argument is that God exists, we should note that the conception of God that emerges from that argument is a relatively thin one. It is what Martin Heidegger called the "god of metaphysics", the "ground of all being". Heidegger himself complained that the god of metaphysics is not one before whom we can sing and dance, the idea being that if our concept of God is *simply* the concept of something that is posited to explain the existence of all other beings, the God we have thus conceived is not one that inspires passionate devotion. But, of course, virtually nobody wielding the Cosmological Argument for the existence of God would think that the concept of God is exhausted by the concept of a first cause; nor would they think that the Cosmological Argument is our only source of insight into what God might be like. To flesh out the concept of God, however, would take us well into the realm of theology, and thus far beyond the proper scope of a book primarily devoted to the basics of metaphysics.

FURTHER READING

On the topic of substance generally, three important works are Michael J. Loux, *Substance and Attribute: A Study in Ontology* (Dordrecht: D. Reidel, 1978), Joshua Hoffman and Gary S. Rosenkrantz, *Substance: Its Nature and Existence* (New York: Routledge, 1997), and Kathrin Koslicki, *Form, Matter, Substance* (New York: Oxford University Press, 2018). Robert Pasnau's *Metaphysical Themes 1274–1671* (Oxford: Oxford University Press, 2011) is another excellent resource. (It is particularly useful on the topic of substance; but it provides a very helpful historical perspective on other topics treated in this book as well.)

On the bundle theory, Max Black's paper, "The Identity of Indiscernibles", *Mind* 61 (1952): 153–164, is famous for pushing the objection that bundle theorists cannot admit the possibility of duplication. Two important discussions of Black's article are John O'Leary Hawthorne and Jan A. Cover, "A World of Universals", *Philosophical Studies* 91 (1998): 205–219, and Dean Zimmerman, "Distinct Indiscernibles and the Bundle Theory", *Mind* 106 (1997): 305–309.

For several different versions of the bundle theory, as well as the main objections discussed in this chapter, see James van Cleve, "Three Versions of the Bundle Theory", *Philosophical Studies* 47 (1985): 95–107. I said that L. A. Paul was among those pushing the bundle theory in the direction of saying that everything is a property. For her own developments of that theory, see L. A. Paul, "Logical Parts", *Noûs* 36, no. 4 (2002): 578–596, and "Building the World from Its Fundamental Constituents", *Philosophical Studies* 158, no. 2 (2012): 221–256.

On bare particular theory, Edwin Allaire, "Bare Particulars", *Philosophical Studies* 14 (1963): 1–7, is a classic article. See also Theodore Sider's more recent "Bare Particulars", *Philosophical Perspectives* 20 (2006): 387–397, and Andrew M. Bailey, "No Bare Particulars", *Philosophical Studies* 158, no. 1 (2012): 31–41. Although Bailey opposes the bare particular theory, his paper is an excellent resource for those looking for references to others who defend the view.

My own understanding of the views of Aristotle and Aquinas, and of hylomorphism in general, is heavily influenced by Michael Loux and Jeffrey Brower. In particular, see Michael Loux, *Primary Ousia* (Ithaca, NY: Cornell University Press, 1991) and Jeffrey Brower, *Aquinas on Material Objects* (Oxford: Oxford University Press, 2013). For contemporary defenses of hylomorphism, see (in addition to the books by Brower and Koslicki that I have already cited) Kit Fine, "Things and Their Parts", *Midwest Studies in Philosophy* 23, no. 1 (1999): 61–74; Kathrin Koslicki, *The Structure of Objects* (New York: Oxford University Press, 2008); and Michael Rea, "Hylomorphism Reconditioned", *Philosophical Perspectives* 25 (2011): 341–358.

On the cosmological argument mentioned at the end of this chapter, I recommend, for starters, the set of readings in Michael Rea and Louis Pojman (eds), *Philosophy of Religion: An Anthology*, 7th ed. (Boston, MA: Cengage, 2013).

8

THINGS AND THEIR PARTS

Imagine placing a toy van in your *Radio-Flyer* wagon. Have you created a new object—a van-in-wagon? Most of us will say no. There are no van-in-wagons; just *vans* and *wagons*. Why? Because simply putting a van in a wagon isn't a genuine way of *unifying* two objects. It is not a way of making *one* object out of *two*. A toy van and a wagon together do not compose anything, or so we normally think. By contrast, consider a process that manages to arrange a bunch of particles human-wise. Imagine a machine that can do this—it can take numerous particles from the nearby atmosphere and put them in exactly the same spatial and causal relations as the particles that compose a human body. (The transporter rooms in the various incarnations of *Star Trek* do something very much like this.) This sort of process, it seems, *is* unifying. Arranging particles human-wise creates a genuine unity: it creates one additional thing where previously we just had a bunch of particles. It gets the particles to compose something whereas previously they did not.

Composition is a puzzling phenomenon. Sometimes when we place objects next to one another (think of a bouquet of flowers) or fasten them together (think of a wall built of brick and mortar) we think we bring something new into existence; sometimes we don't (think of books side-by-side on a shelf, or a piece of paper that has

become stuck to a countertop thanks to some spilled milk that has dried). Likewise, **material constitution**—the relationship between an object and the mass of matter that makes it up—is also deeply puzzling. Consider, for example, a piece of playdough that is molded into a statue of Gumby. Gumby doesn't survive when the playdough is squished back into the shape of a ball, but the piece of playdough does survive. So it seems they must be different things. But, at the same time, it seems like they *can't* be different things, for in that case we would have two things occupying the same place at the same time. So which is it? Is Gumby something distinct from the mass of playdough that constitutes it, or are "they" one and the same thing?

Questions about how things and their (alleged) parts relate to one another have been a longstanding staple of both contemporary and historical metaphysics. In this chapter, we begin by exploring the topic of composition, and then we turn our attention to puzzles about material constitution.

COMPOSITION

In his widely influential book *Material Beings*, Peter van Inwagen raised the following question: What, in general, does one have to do, and what would suffice for, getting some objects to compose some other object? He called this question the "special composition question" (henceforth, SCQ for short). Some answers can be dismissed right away: Gluing things together, for example, is not sufficient for getting objects to compose something. Glue two people together and, it seems, we do not thereby create a new object. In fact, according to van Inwagen, no particular kind of bonding is sufficient. Bond two human beings together in whatever way you like; that by itself will not create a new object. Thus, much of the focus in the vast literature that has been spawned by the SCQ has been on two "extreme" answers: *universalism*, according to which any objects whatsoever compose something; and *nihilism*, according to which there are no composite objects at all.

Given the commonsense assumption that vans, wagons, fish, moons, and so on all count as genuine composite objects, universalism implies that, for example, there are also objects like the van-in-wagon mentioned earlier, as well as an object composed of your nose and the moon, or of all the fish in the sea, and so on. Many

philosophers find this view counterintuitive. Just as it doesn't seem that two human beings glued together compose something, so too it doesn't seem that all the fish in the sea, or your nose plus the moon, compose anything. All the fish in the sea seem more like a mere plurality—just a bunch of fish, and not a single, unified material object. Likewise for your nose and the moon.

In addition to having counterintuitive implications for what objects exist, universalism as it is often developed also seems to be inconsistent with obvious empirical facts. In his own argument against universalism, van Inwagen cites the fact that human bodies undergo complete change of their smallest parts every ten years or so. Let *the ps* be the particles that composed my body fifteen years ago. Right now, the *ps* do not compose my body (since, again, bodies undergo complete replacement of their smallest parts roughly every decade). But if universalism is true, the *ps* still compose something—presumably a scattered cloud of particles. The question, however, is just *when* they went from composing a human body to composing something else. Consider a time *t* when the *ps* definitely composed my body. Those particles, of course, are constantly in motion. Presumably some degree of motion is consistent with them continuing to compose my body; but as soon as one of them moves far enough away from my body, there is pressure to say that the *ps* no longer compose my body but instead compose something else (something very much like my body, plus a stray particle). Indeed, it looks as if we will have to say that this "something else" is a distinct object that very significantly *overlaps* my body, since, if the *ps* no longer compose my body but my body still exists, it follows that some other group of particles (namely, the *ps* minus that stray particle) composes my body. But does that picture really make sense? Many universalists think it doesn't—if universalism is true, then (for any *ps* whatsoever), the *ps* always compose something, and, furthermore, they never compose more than one thing either at the same time or in succession. If this is true, then universalism is inconsistent with the supposedly "obvious empirical fact" that human bodies and other things can and normally do survive changes in their parts.

I think there are replies that can be made to this argument; and, in fact, universalism is my own preferred answer to van Inwagen's SCQ. But let us leave that view aside for now and turn to other answers.

Nihilism lies at the other extreme on the continuum, maintaining that composition *never* occurs. This, too, looks to be inconsistent with obvious empirical facts. Look around you. Isn't it plain as day that there are all manner of composite material objects. Do we really want to say that tables, chairs, houses, computers, cars, trees, our own bodies and the bodies of our friends and pets and so on do not exist? Of course, nihilists are sensitive to this concern, and they certainly don't intend to deny what is plain to the naked eye. But, they argue, what is plain to the naked eye is just that the regions we take to be occupied by tables, chairs, human bodies, and so on are filled by matter arranged (spatially and causally) table-wise, chair-wise, human-body-wise, and so on. On their view, however, it is *not* plain to the naked eye that the particles filling those regions actually compose something.

If this reply seems implausible, consider a different case: a bouquet of flowers. Take a bouquet of flowers up to the checkout counter at your local market and you will pay one price for that one item. If you are in an express lane, eight items or less, you will not be in violation of the rules if your bouquet includes a dozen roses. But is it *obvious* that the flowers compose a single, unified object? After all, you also won't be held in violation of the express lane's rules if you show up there with a bag of twelve tomatoes; but it is hardly an obvious empirical fact that a dozen tomatoes compose a single object. So too, says the nihilist, in the case of the particles that (allegedly) compose a table, a chair, or a human body. If anything tells us that any of those groups of particles compose something, it is philosophical theory, not visual experience. Still, if the nihilist is right, one straightforward consequence is that *nothing is destroyed* if you blow up a building, dismember a body, cut down a tree, etc. And this consequence, for many philosophers, is too hard a pill to swallow.

So far we have been looking only at extreme, all-or-nothing answers to the SCQ. Van Inwagen's own answer, however, lies in between the extremes. According to van Inwagen, composition occurs whenever you arrange things so that their activity constitutes a *life*. Part of what motivates this answer is the fact that it is hard to find a plausible moderate answer to the SCQ that will recognize the existence of *all and only* the composites that so-called common sense recognizes: organisms, artifacts, and natural bodies like rocks,

mountains, planets, and so on. (If you doubt this, just try to come up with your own moderate answer. You will almost certainly find that it either recognizes the existence of composites you don't believe in, or denies the existence of composites you do believe in.) But the more important motivation, the reason that keeps van Inwagen from going to the extreme of endorsing nihilism, is the following collection of (to his mind) non-negotiable claims, some of which I'll express here, as he does, in the first person: I exist; I think; thinking requires a thinker; and *if* I exist, I am a material object, an organism. If these claims are true then there is at least one composite object—namely, me. So nihilism is false. The question, then, is what feature of the particles that compose me "makes" them compose something. Van Inwagen thinks that the most natural, least arbitrary answer is that *their activity constitutes a life*. And thus we get his answer to the SCQ.

Why should we believe that thinking *requires* a thinker? According to van Inwagen, whereas the activities of artifacts like shelves, automobiles, and so on are plausibly thought of as "disguised cooperative" activities, thinking is different. On his view, it is easy to see how particles might work together without composing anything to hold up books or to move a human being down the road, but it is not easy to see how they might work together to think without composing anything. This seems right; but it is worth noting that the activities of sophisticated computers are relevantly similar to the activities of thinking organisms. If, as van Inwagen thinks, the particles that (allegedly) compose a computer don't need to compose anything in order to do their business, it is hard to see why the particles that (allegedly) compose me would *have* to compose something in order to do engage in the very similar, albeit more sophisticated, business of human thought. But if this is right, then we are "forced" to believe in organisms only if we are also forced to believe in computers; and so if it is acceptable to endorse an answer to the SCQ that eliminates artifacts altogether, it should also be acceptable to endorse full-blown nihilism. For this reason, then, van Inwagen's argument in support of his answer seems unconvincing.

How shall we make progress on the SCQ? One approach is to return again to the hylomorphic theory of substance discussed in Chapter 7. Hylomorphism does not specify fully informative

conditions under which composition occurs. To do that, one would have to identify the conditions under which a mode of arrangement rises to the level of being a (substantial or non-substantial) form; and hylomorphism as such does not help us to identify such conditions. But it can at least provide a framework for thinking about the SCQ that reveals what some of the salient issues might be.

Hylomorphism alone among the theories of substance discussed in Chapter 7 allows that the primary substances in the world are composite objects that undergo change and development. Furthermore, the hylomorphist separates the substances from the non-substances precisely on the basis of differences in the mode of arrangement displayed by the (alleged) parts of a thing. Those modes of arrangement that impose *genuine unity* upon things join with matter to generate substances; those that do not, do not. Thus, hylomorphism implies that wherever we find a mode of arrangement that imposes genuine unity upon things, we find a composite object. So we have a sufficient condition, even if not a maximally informative one, for when composition occurs.

For the hylomorphist, then, the first decision point in addressing the SCQ is to ask whether there are substantial forms. It would be strange, though not impossible, to be a hylomorphist and to believe that no modes of arrangement whatsoever count as substantial forms. Thus, the likely answer to the question whether there are substantial forms is "yes". Nihilism is ruled out.

Another decision point is to ask whether all forms are substantial forms. Suppose we say yes. It is clear that not *every* way of arranging objects is going to count as a substantial form. My children arrange their toys in my living room every day in ways that display no functional unity, nor any other obvious form of unity. If the toys together did compose something, what they compose would have no internal principle of change and development. In short, it would have no substantial form. Indeed, if we said that even the toys in my living room *do* compose an object with a substantial form, we would likely lose all basis for saying that there is a distinction to be drawn between genuine unity and merely apparent unity. The substance/non-substance distinction would become meaningless, and so there would be no point in characterizing *any* forms as substantial forms. So if we say that *all* forms are substantial forms, we should also say that the toys in

my living room do not compose something. Universalism is therefore ruled out.

Universalism is *not* ruled out, however, if we allow that there are non-substantial forms. In that case, we might say that some attribute collectively exemplified by the toys is playing the role of form, albeit rather poorly, or only in an extended sense. That might suffice for the toys to compose an object, even if not a substance. And we might say the same for *any* attribute collectively exemplified by *any* objects whatsoever. If we do say this, then we have opted for universalism after all. In saying this, however, we will still want to draw distinctions between modes of arranging objects, identifying some as sufficiently rich and unifying to count as substantial forms and identifying others as falling short of that threshold, even if they do nevertheless impose some relatively weaker sort of unity upon things. (We will want to say that non-substantial forms genuinely unify their parts. The difference between substantial and non-substantial forms will just be in the degree and kind of unity imposed.) In saying all of this, then, we can, even if we endorse universalism, do justice to the intuitions of those who prefer moderate answers to the SCQ. We can say that those who endorse moderate answers—like van Inwagen's answer, according to which only living organisms count as composites—have incorrectly identified the substance/non-substance distinction with the object/mere plurality distinction. This doesn't resolve the dispute, but it clarifies a significant issue on which the dispute depends.

A final decision point concerns the criteria for genuine unity. What must a mode of arranging objects be like in order to impose genuine unity upon objects? (If we believe in both substantial and non-substantial forms, we might further ask what a mode of arranging objects must be like in order to impose *enough* unity upon objects to generate a substance.) Must it be a mode of *functional* arrangement? Must it be a mode of arrangement that, given the laws of nature, guarantees that the object whose form it is will change and develop in predictable ways all of its own accord in something like the way that living organisms do? Must it be a mode of arrangement that guarantees that the object whose form it is will be self-maintaining in some sense? These are difficult questions—no less difficult, really, than the SCQ itself.

The fact that we still face such difficult questions, however, does not mean that hylomorphism makes no progress with the SCQ. As we have seen, endorsing hylomorphism gives us at least two clear decision points that will help us to decide between extreme answers to the SCQ. Moreover, reframing the SCQ, as hylomorphism encourages us to do, as a question about what it takes to unify some objects opens up avenues of exploration that might otherwise not have occurred to us. There is, for example, a sizeable philosophical literature on functions that can be brought to bear on the question of what it takes to functionally organize objects. Doing so might shed light on the question whether functional organization is really any more unifying than other modes of arrangement. It might also shed light on the question whether sharp lines can be drawn between things that are functionally organized and things that are not. All of this can help the hylomorphist to flesh out their understanding of form and, in so doing, to craft a more fully informative answer to the SCQ.

MATERIAL CONSTITUTION

Suppose you loan money to a friend, a substantial sum. Months later, weeks past the deadline for repayment, you contact your friend and ask for your money. Your friend offers instead a philosophical argument:

> Look at me: I'm just a collection of particles. You don't think there's something here *in addition* to the collection of particles, do you? I mean, it's not as if, in addition to the 135 pound collection of particles there is *also* a 135 pound person, right? So there is a collection here, and I am that collection. But, now think about it for a moment. I am obviously a *different* collection of particles from the one who borrowed money from you. For example, *this* collection of particles includes some of the stuff that composed my breakfast this morning; *that other* collection didn't include that stuff. So I am a different collection of particles from the one who borrowed the money. A human person *just is* a collection of particles. So I am a different person from the one who borrowed that money. So I don't owe you any money.

Obviously something has gone wrong. But what? How is the debt-or's argument flawed?

The debtor has raised a puzzle about material constitution. Material constitution occurs whenever an object *a* and an object *b* (in this case, the debtor and the collection of particles that constitutes them) share all of the same parts at the same time. The phenomenon is puzzling because it is not at all clear what the relationship is between the relevant *a* and *b*. In particular, it is not clear whether a = b. The debtor assumes that they are identical with the collection of particles that now constitutes them and concludes that they are therefore not the same person as the one who earlier borrowed some money (because that person was identical to a different set of particles). No doubt we want to deny this. But what will we say instead? If we say that they are *not* identical to the collection of particles that constitutes them, do we say that there is a 135 pound person and a distinct 135 pound collection of particles occupying *exactly the same place at the same time*? If not, then what else can we say? We seem to have no good alternatives.

There are many different kinds of puzzles about material constitution in the philosophical literature; but all of them present us with scenarios in which it appears that an object *a* and an object *b* share all of the same parts but have different modal properties. (Modal properties, again, are properties like *being able to survive being squashed* or *being possibly made of wood*.) In what follows, I will sometimes also speak of **persistence conditions**. Since persistence conditions are facts about what changes an object *can* or *cannot* survive, they too are modal properties. The fundamental problem that they all raise is known as "the problem of material constitution".

I will discuss solutions shortly, but first it is important to get several of the different puzzles out onto the table. The Debtor's Paradox, presented above, is one of them. In what follows, I will discuss three others: The Ship of Theseus, Body-Minus and Lumpl/Goliath. Familiarity with these three puzzles and their possible solutions is a gateway into a wide variety of core topics in contemporary metaphysics.

THE SHIP OF THESEUS

Consider a ship—the Ship of Theseus. At the beginning of its career, the ship is made entirely of wooden planks. The ship sails

the same route for many decades and is "preserved" in the following way: Whenever one of the wooden planks wears out, it is discarded and replaced by an aluminum one. Eventually the time comes when all of the wooden planks have been replaced by aluminum ones. One day, however, an historian decides to gather all of the discarded planks and rebuild them in their original form. As a result of their work, each plank has the same position that it did in the original ship. The historian sells the ship to the local museum, and the curator then boasts that the museum has on display the Ship of Theseus. The crew of the aluminum ship, however, is outraged: "*We* are sailing the Ship of Theseus and have been for many years. The Ship of Theseus is here on the water, not there in your museum!" Who is right? Which ship is the Ship of Theseus?

BODY-MINUS

Consider Tibbles, a perfectly ordinary cat sitting happily in the sun, cleaning her tail. Now imagine at some later time an unfortunate accident befalls her, resulting in the annihilation of her tail. Tibbles is understandably distressed. But we should be also. To see why, let us use "Body-minus" to name that part of Tibbles that includes all of her except for her tail—call it her *torso*. Prior to the accident, Tibbles is distinct from her torso. It is a **proper part** of Tibbles (i.e., the torso does not completely coincide with Tibbles). After the accident, Tibbles survives. Annihilating a tail doesn't normally destroy a cat. But Body-minus also survives. Annihilating a tail doesn't destroy a torso. But now we have a problem. On the one hand, we want to say that, after the accident, Tibbles is identical to Body-minus. Otherwise we would have two things in the same place at once. But Body-minus and Tibbles have different modal properties. Body-minus can exist as a mere torso but Tibbles cannot. Tibbles can be a cat with a tail; Body-minus cannot. So it seems that Tibbles cannot be identical to Body-minus after all.

LUMPL/GOLIATH

This puzzle was introduced by Allan Gibbard. It is similar to the puzzle about Gumby that I quickly stated in the introduction to this chapter, except it is crafted in such a way as to preclude certain replies that are

available in the Gumby case, but not in *every* case of material constitution. A sculptor sets out to create a statue of Goliath, but they do so in the following way: they sculpt Goliath's upper body from one piece of clay, his lower body from another, and then they finish the statue by sticking the two pieces of clay together. In joining the two pieces of clay the sculptor simultaneously brings into existence a *new* piece of clay (call it "Lumpl") and a completed statue (call it "Goliath"). They allow the clay to harden, but then a day later, dissatisfied with their work, they smash the statue, thereby obliterating both Lumpl and Goliath. The question is whether Goliath is identical to Lumpl. There seems to be no good answer to this question. An affirmative answer is implausible since Lumpl and Goliath seem to have different modal properties. For example, (supposing the clay is still wet) Lumpl would survive if the piece of clay were reshaped in the form of a vase, whereas Goliath would not. And Goliath would survive if its left finger were annihilated, whereas Lumpl would not. But a negative answer is also implausible, at least initially, since such an answer appears to commit us to the conclusion that two distinct objects (a lump and a statue) can fully occupy the same region of spacetime.

All of these puzzles have roots in antiquity; and there are endless variations on each. But why think that they are all puzzles about material constitution? What are the relevant objects *a* and *b* in each puzzle that (allegedly) share all of the same parts and yet have different modal properties? The answer is straightforward in three of the puzzles. In the Debtor's Paradox, the relevant objects are the debtor and the collection of particles that constitutes them. Intuitively, the persistence conditions of persons and mere collections of particles are different: removing particles from a collection leaves one with a *new* collection, whereas removing a few particles from a person doesn't seem to leave one with a new person (nor does it seem to destroy a person). Hence the difference in modal properties. In the Lumpl/Goliath puzzle, the objects are Lumpl and Goliath. Intuitively, lumps of clay and statues have different persistence conditions: The former but not the latter can survive being squashed. Hence they too have different modal properties. In the Body-minus puzzle, the objects are Tibbles and Body-minus; and I have already explained how it is that they have different modal properties.

The Ship of Theseus Puzzle is more complicated. Initially, it seems that the puzzle presents us not with two objects but at least *three*: the original ship, the aluminum ship, and the historian's ship (i.e., the ship that has been rebuilt from the original planks). Moreover, at the end of the story it is not at all clear which (if any) of these objects constitute one another. Obviously the historian's ship and the aluminum one do not constitute one another. But does either constitute the original ship? That question seems equivalent to the one already under dispute. For if either of the two ships constitutes the original, it will do so only because it is *identical* with the original. Thus, it is hard to see how this puzzle presents us with an object *a* and an object *b* that share all of the same parts but appear for some reason to be *distinct* from one another.

But suppose we focus our attention not on the end of the story but on the beginning. At the beginning of the story, what kind of thing do the original planks compose? A ship, obviously. But what kind of ship? We start with the intuition that they compose a ship that can survive complete part replacement, and this is what tempts us to identify the aluminum ship with the original. But during the story, another of our intuitions is uncovered. It turns out that we are also inclined to think that, at the beginning of the story, the planks compose a ship that *cannot* survive complete part replacement. This intuition is what tempts us to identify the historian's ship with the Ship of Theseus. So, in other words, the puzzle arises because we have some reason to think that two ships coincide at the beginning—one which can survive complete part replacement and one which cannot. By the end of the story, those two ships have split apart; one is made of aluminum, the other sits in a museum. As with any puzzle about material constitution, we can just accept the conclusion that there were two ships at the beginning where we thought there was only one. Or we can insist that, at the beginning of the story, either the aluminum ship or the historian's ship did not exist, thus identifying the remaining one with the Ship of Theseus. Or we can look for some other solution to the puzzle.

What, then, are our options? How do we solve these puzzles? Speaking quite generally, each puzzle presents us with a choice among four solutions.

OPTION 1: DENY THE EXISTENCE OF OBJECTS BELONGING TO THE KINDS THAT ARE CENTRAL TO THE PUZZLE

Each of the puzzles assumes the existence of at least two objects, each belonging to a different kind. The way these puzzles normally go, we start with a familiar kind of composite object—e.g., ship, statue, cat. We are then invited to believe in the existence of another kind of object—e.g., mere torso, mere collection of particles, etc. Associated with each kind are different modal properties—e.g., statues can't survive squashing, lumps can; cats can have tails, torsos can't; etc. This is what generates the puzzle. Thus, each of the puzzles can be solved by denying its existence assumptions. If there are no statues, or no mere lumps, then the Lumpl/Goliath puzzle is solved. If there are no torsos, the Body-minus puzzle is solved. If there are no composite objects at all, then all of the puzzles are solved. One of the attractions of nihilism, discussed in the previous section, and of answers to the SCQ like van Inwagen's that rule out the existence of artifacts, is that they provide just this kind of solution to the problem of material constitution.

OPTION 2: DENY THE PROBLEMATIC MODAL PROPERTIES

As I have just noted, in every constitution puzzle, different modal properties are associated with each of the kinds mentioned in the puzzle. So, for concreteness, consider again the Lumpl/Goliath puzzle. In raising the puzzle, we assume that lumps have one set of modal properties, statues another. But we can contest this assumption. So, for example, we might insist that at least some lumps—namely those that are also statues—cannot survive being squashed. We might insist that at least some collections of particles—namely, those that are human beings—have the modal properties of persons rather than the modal properties of mere collections. And so on.

We can put this solution more generally as follows. Suppose you are confronted with an object *a* and an object *b* that share all of the same parts but belong to two different kinds (K1 and K2); and suppose that the modal properties associated with K1 (e.g., *being able to survive squashing*) are incompatible with those associated with K2 (e.g., *being unable to survive squashing*). What you should say is the

following. $a = b$, and that object belongs to both kinds. But not every object can have all of the modal properties associated with all of the kinds to which it belongs. Rather, every object has a *dominant kind*. An object's dominant kind is the one you'd refer to if you were in a position to give, for the object, the correct answer to the question, "What is the nature of that object?" So, in the case at hand, perhaps K1 is the dominant kind; perhaps K2 is the dominant kind; or perhaps neither is. Regardless, however, it won't be true that the object in question has the (incompatible) modal properties associated with *both* kinds.

This is an attractive solution, but it carries at least one unattractive cost. Suppose you shape a piece of clay into a statue. The piece you start with is a mere lump. The statue is also a lump; but (we might suppose) its dominant kind is statue. Since the initial lump was *by nature* a mere lump and the statue is not, it follows that the statue is not identical to the lump you started with. Shaping the lump destroyed it and replaced it with a statue. Similar results arise in the other constitution puzzles. This, then, is our cost—the present solution implies that what appear not to be destructive changes *are* destructive changes. Some philosophers find this to be highly counterintuitive.

OPTION 3: DENY THAT IDENTITY IS NECESSARY

Each of the puzzles tacitly assumes that identity is necessary. That is, each puzzle assumes that, for any objects x and y, if x is identical to y, then *x has to be* identical to y. (It follows from this assumption that if x is distinct from y, then *x has to be* distinct from y.) Denying this assumption—saying that identity is **contingent**—therefore suffices to solve all of the puzzles at once. The Debtor's Paradox would go away if only we could say that, although the debtor is *now* identical to the collection of particles arguing about the debt, they *were* identical to the collection that contracted the debt. The Body-minus puzzle would go away if only we could say that Body-minus *is now* (after the tail-removal) identical to Tibbles, but *was not* identical to Tibbles when it was a mere torso. The Lumpl/Goliath puzzle would go away if only we could say that Lumpl *is* identical to Goliath but might not have been. The Ship of Theseus puzzle would go away if

only we could say that Ship A *was* identical to Ship B but is no longer. Similarly for any other puzzle about material constitution.

The cost of this solution is just the fact that it is extremely counterintuitive. How could a thing possibly be distinct from itself? That said, this solution is not universally rejected. But its defenders tend to be few. Those who do defend it typically do so by adopting a somewhat unusual view about what it would mean to affirm the sentence, "Lumpl could have been distinct from Goliath" given that, in fact, Lumpl = Goliath. The most well-known version of this unusual view is counterpart theory (discussed already in Chapter 4).

OPTION 4: DENY THE ASSUMPTION THAT IF X AND Y SHARE ALL OF THE SAME MATERIAL PARTS, THEN X IS IDENTICAL TO Y

This option, of course, has been on the table from the very beginning of our discussion. None of the puzzles would be very puzzling if we did not have the strong intuition that two distinct things can never occupy exactly the same region of spacetime at the same time or share all of the same material parts. But one might think that precisely what the puzzles show us is that this deeply held intuition is false.

The most obvious way to deny the assumption that if x and y share all of the same material parts then $x = y$ is simply to admit that two distinct material objects—two things that are not in any meaningful sense the *same* material object—can occupy exactly the same region of space at once. But another way of denying this assumption is to endorse an "Aristotelian" solution to the problem of material constitution.

In the previous section, we saw how hylomorphism provides a framework for thinking about the SCQ. Together with what seem to me to be fairly intuitive supplementary claims, it also provides the resources for saying that objects that constitute one another are the *same material object* despite being *numerically distinct things.* Consider again the Lumpl/Goliath puzzle (and let us suppose for now that *statue* is a substance kind and *lump* is not). In the region occupied by Goliath there appear to be two distinct material objects: a lump and a statue. Note, however, that within a hylomorphic framework the lump and the statue can be seen as standing in an interesting kind of

relationship with one another: Either one is the matter for the other (i.e., perhaps the statue is just a compound of the lump and the form *being a statue*) or they share the same matter (i.e., the same underlying plurality, or the same prime matter, bears a substantial form—*being a statue*—and one of those secondary, non-substantial forms mentioned in connection with "the third option" discussed earlier—*being a lump*). We can run these two options together by saying that two things share all of the same matter in common if, and only if, either one is matter for the other or the matter for one is also the matter for the other.

The question, then, is what we shall say about objects like this that share all of the same matter in common. Notice, first of all, that only one of the objects is a substance: the statue. So right away we preserve the intuition that there is just *one* primary subject of predication in the region. However, we also preserve the intuition that the lump is a real thing, and is something different from the statue. Still, we face the question of how many *material objects* there are in the region. Here, Aristotle has something interesting to say. In cases like this one, the Aristotelian view seems to be that the statue and the lump are *one in number* but not *one in being*. That is, they are different things, but are to be counted as one material object. The statue and the lump, then, are *numerically the same* material object. Goliath *is* Lumpl, and there is just one material object in the region occupied by it. But they are nevertheless distinct *hylomorphic compounds*.

Initially this seems completely baffling. How can *two hylomorphic compounds* be *one and the same material object*? The answer comes in two parts. First, we note that this solution presupposes the following (controversial) principle of *relative sameness*:

(RS)   States of affairs of the following sort are possible: x is an F, y is an F, x is a G, y is a G, x is the same F as y, but x is not the same G as y.

In the case at hand, Lumpl and Goliath provide an instance of RS. Each is a material object, and each is a hylomorphic compound; and Lumpl is the same material object as Goliath but not the same hylomorphic compound. How can this be? Here we come to the

second part. What it *is* for one thing to be the same material object as another is just for the one thing and the other to share the same matter in common. That is just part of the concept of being a material object—we count one material object wherever we have some matter with one or more forms imposed upon it.

That last claim should be completely intuitive. It is, I think, commitment to that last claim about the conditions under which we count one material object that makes us recoil at the thought of saying that Lumpl and Goliath are *two* material objects. But notice that we do not recoil at the thought of saying that two *things* occupy the same place at the same time. Two events can easily occupy the same place at the same time—e.g., the event of your sitting and the event of your thinking about philosophy. Hylomorphism gives us the resources to respect these basic intuitions. The fact that substances are compounds of matter and form, together with the fact that the same matter can bear multiple forms (even if some of the forms are not *substantial* forms), gives rise to cases where we want to count *two* things (two hylomorphic compounds) but just *one* material object (since, again, we count one material object wherever we have some matter bearing one or more forms).

FURTHER READING

On the topic of composition, Peter van Inwagen's *Material Beings* (Ithaca, NY: Cornell University Press, 1990) is a must-read. In defense of nihilism, see Theodore Sider's "Against Parthood", *Oxford Studies in Metaphysics* 8 (2013): 237–293. For an influential defense of a view similar to (but not quite the same as) van Inwagen's, see Trenton Merricks, *Objects and Persons* (Oxford: Oxford University Press, 2001). For my own defense of universalism, see "In Defense of Mereological Universalism", *Philosophy and Phenomenological Research* 58 (1998): 347–370. Several of the essays in the *Metametaphysics* volume cited at the end of Chapter 1, as well as Karen Bennett's *Making Things Up* (New York: Oxford University Press, 2017) provide valuable recent treatments of the topic of composition and related issues, and readers can get a good sense from these works of the overall shape of the discussion of this topic in the literature over the past three decades.

On the subject of material constitution, a good place to start is Michael Rea (ed.), *Material Constitution: A Reader* (Lanham, MD: Rowman & Littlefield, 1997). Allan Gibbard's paper, "Contingent Identity", is among the articles reprinted

there. The topic of material constitution is also discussed at some length in Theodore Sider's *Four-Dimensionalism* (Oxford: Oxford University Press, 2001) and Katherine Hawley's *How Things Persist* (Oxford: Clarendon Press, 2001). Finally, interested readers will benefit from looking at Ryan Wasserman's valuable discussion in "The Constitution Question", *Noûs* 38 (2004): 693–710.

9

CHANGE AND IDENTITY

Central to our commonsense view of the world is the idea that familiar objects can persist through a wide variety of changes. People grow up; trees drop their leaves in the fall and have them restored in the spring; caterpillars become butterflies. But no matter how obvious—indeed, trite—such claims might seem, the phenomenon of change is fraught with serious difficulties.

In this chapter, we take up four of the most important and apparently intractable arguments against change and identity over time that have occupied philosophers for the past two millennia. We begin with an argument against the reality of change that is due to the Presocratic Greek philosopher Parmenides, followed by three arguments against the reality of motion that were given by his famous disciple, Zeno. The arguments of these philosophers exerted powerful influence on the shape of Western philosophy. Many in their wake were captivated by the idea that the changing world of sensory experience is but a mere shadow or reflection of a more fundamental, more deeply real world of unchanging things. Next, we discuss a contemporary argument for the thesis claim that no object can change its **intrinsic properties**. Finally, we focus on the special problems that arise when we consider how *persons* can survive through change.

PARMENIDES ON CHANGE

Parmenides of Elea, who lived during the 5th and 6th centuries BCE, is generally regarded as the most important and influential of the Presocratic philosophers. His philosophical views were articulated in a poem, large fragments of which have been preserved through the writings of other philosophers. In the poem, Parmenides describes his journey to meet a goddess who reveals to him two paths of inquiry: the Way of Truth and the Way of Mortal Opinion. Those who follow the latter fall into all manner of "erroneous" beliefs that most of us would regard as plain common sense: the world is filled with many different things, these things come into existence and pass out of existence, they change, they move, and so on. On the Way of Truth, however one finds arguments for the conclusion that all of these commonsense beliefs are mere illusions. Parmenides endorsed a version of what is now known as "Eleatic Monism", the view that there exists exactly one thing that is ungenerated (it did not come into existence), indestructible (it cannot go out of existence), does not change or move and is undivided in time and space (it has no spatial gaps or holes within itself, and there are no gaps between the times at which it exists either). Parmenides's monistic worldview was developed and defended by his followers, especially Zeno and Melissus. Our concern in this section is with their arguments against change and motion.

Parmenides held that whatever exists is wholly unchanging. His defense of this view rests on the idea that *non*existence is unintelligible. According to Parmenides, it is impossible to talk about or even to think about what does not exist. The reason, presumably, is that what does not exist is not available to be an object of thought or discourse. Thus, what does not exist cannot be an object of inquiry; nor can it be referred to by the subject term of any sentence. But then it looks as if nonexistence claims—claims, like "Santa Claus does not exist"—cannot possibly be true. This is the first major step in his argument against change.

The next step should be easy to guess: Generation and destruction involve nonexistence. Suppose that Parmenides himself came into existence at some time. This supposition obviously implies that there once was a time when Parmenides did not exist. But, if Parmenides is

right, the proposition that *Parmenides does not exist* cannot possibly be true. So the supposition that there once was a time when Parmenides did not exist must be false. Therefore, Parmenides is ungenerated. Similar reasoning rules out the possibility of Parmenides passing out of existence; thus, he cannot be destroyed.

Parmenides seems to have thought that all change involves generation and destruction. Suppose that Parmenides stands up. This change in Parmenides occurs when the event *Parmenides' standing* is generated and the event *Parmenides sitting* is destroyed. Likewise, Parmenides changes from not cheering to cheering when the event *Parmenides' cheering* is generated and the event *Parmenides standing quietly* is destroyed. But generation and destruction are impossible. So these changes cannot possibly occur. Likewise for any change whatsoever.

Parmenides's arguments rest on the thesis that nonexistence claims cannot be true. In Chapter 2, we discussed some of the main reasons one might have for endorsing this thesis, along with the major strategies for resisting it. To recap briefly: Suppose we say that Pegasus does not exist. This seems to imply that *there is something that does not exist*. But, assuming *being* and *existence* are the same thing, this latter claim is self-contradictory. One way to avoid the contradiction is to admit that Pegasus exists. Another is to sever the link between being and existence. A third is to reinterpret the claim that Pegasus does not exist as saying something like *nothing Pegasizes*. There are other solutions in the contemporary literature as well, most exploring the semantics of "empty names" (names that seem not to refer to anything). By and large, these solutions can be seen as further instances of the "reinterpretation" strategy. Importantly, however, none of these solutions has immediate intuitive appeal. The basic problem that Parmenides has handed down to us is a formidable one.

Zeno of Elea lived in the 5th century BCE and has passed on to us a variety of puzzles aimed at defending various theses of Eleatic monism. Here we shall focus on three of his paradoxes aimed at demonstrating the impossibility of motion. Zeno did not deny the *appearance* of motion. His point was simply that the appearance is misleading—ultimate reality does not include motion.

Whereas Parmenides's arguments traded on the unintelligibility of nonexistence, Zeno's exploit the puzzling nature of the infinite.

Zeno assumes that space and time are continuous—they are composed of unextended points or, in the case of time, instants, and between any two points or instants there are infinitely many other points or instants. The main idea underlying the first two paradoxes is that motion from point A to point B requires that the mover first touch the midpoint between A and B. The main idea underlying the third paradox is that nothing is in motion at an instant.

With this in mind, consider first the Traveler Paradox. Suppose that in a finite period of time, Frodo travels a finite distance (from the Shire to Rivendell). Prior to reaching Rivendell, Frodo will have to reach the midpoint between the Shire and Rivendell (Weathertop); and prior to reaching Weathertop, Frodo will have to reach the midpoint between the Shire and Weathertop (somewhere outside of Bree); and … *ad infinitum*. So there will be an infinity of midpoints that Frodo will have to reach before he gets to any particular destination. But it is impossible to reach an infinite number of points in a finite period of time. So, in general, motion is impossible.

One might respond by saying that the argument falsely presupposes that Frodo would have to *touch* every point along his path between the Shire and Rivendell in order to get to Rivendell. Small as he is, Frodo is too big to do that. He won't touch *every* point; he will merely "cross" every point—which, one might think, is less problematic. But we can reframe the problem as follows. Consider a cube sliding across a desk. As the cube slides, so does its two-dimensional front surface. Presumably the two-dimensional surface is of the right size to "touch" literally every point along the distance from one end of the desk to the other. But, of course, the surface can't do that—again, because there are infinitely many points along that distance.

The Paradox of Achilles and the Tortoise is similar. In order for a faster runner to catch a slower runner, the faster runner must first reach the point from which the slower runner started. But by the time Achilles gets to the tortoise's starting point (call it S1), the tortoise will have moved on to a new "starting point" (S2). So now, once he has reached S1, Achilles can catch the tortoise only if he first reaches S2—by which point the tortoise will again have moved on to a new starting point (S3) … and so on, *ad infinitum*.

As with the Traveler Paradox, the Paradox of Achilles and the Tortoise exploits the apparent fact that some tasks are such that, in

order to perform them, one must be able to perform what contemporary philosophers and mathematicians call a *supertask*. In his *Stanford Encyclopedia of Philosophy* article on supertasks, J. P. Laraudogoitia defines a supertask as "an infinite sequence of actions or operations carried out in a finite interval of time". Some believe that supertasks are impossible, and deny the apparent fact that we would have to be able to perform them in order to move from one point to another; others believe that they are possible and provide explanations as to how they might be so.

The third puzzle about motion handed down to us from Zeno is the Paradox of the Arrow. This one trades on the continuity of time rather than the continuity of space. Begin by considering the world at an instant (i.e., a point in time). Nothing moves at an instant—motion takes place over time. Now, consider a flying arrow. Since nothing moves at an instant, it seems that, at every instant at which it exists, the arrow will have to be stationary. But if the arrow is stationary at every instant of its existence, it is not moving—and so it is not flying. Hence our paradox. Unlike the other two paradoxes, this one arises out of the fairly intuitive idea that a dynamic universe cannot be built from merely static instants. Solving the paradox in a satisfying way therefore requires us either to deny that time is continuous or to arrive at an understanding of motion that helps us to see how *being in motion* is a property that something like an arrow can have even at an instant.

THE PROBLEM OF TEMPORARY INTRINSICS

Eleatic arguments against change and motion gripped the ancient world and exerted heavy influence on subsequent metaphysical theorizing. But they are hardly the only puzzles about change with which a metaphysician must contend. Another puzzle—one which has garnered a great deal of attention in the contemporary philosophical literature—is the so-called "problem of temporary intrinsics", a puzzle which raises difficult questions about the possibility of intrinsic change.

Intrinsic properties are attributes that something can have regardless of what the rest of the world is like. For example, *being bipedal* is intrinsic, because having it depends only on how one's own parts are arranged and not on what anything outside of oneself happens to be

like. *Being married* or *being the only biped in the world* are not intrinsic, because having those properties depends to some extent on what the rest of the world is like. Being married requires the existence of someone else, and it requires furthermore that you have a particular sort of relationship to that person. Being the only biped requires the nonexistence of other bipeds.

With this in mind, consider an ordinary case of intrinsic change: Homer is sitting at *t1*, and so at that time has a *bent* shape. He stands up at *t2*, and so at that time has a *straight* shape. We can begin to see the problem by considering a spatial analogy. Think of places like the Notre Dame football stadium, the White House, or the Redondo Beach pier. If Homer is sitting in the Notre Dame football stadium, then Homer is sitting, period. If Homer's sitting is part of an event that we might describe as *everything happening at the Redondo Beach pier*, then Homer is sitting, period. In general, what occurs *at a place* occurs, period. Why should times be any different? But if we say that matters are not different in the case of times, then we will endorse the view that whatever occurs *at a time* occurs, period. So if Homer is bent at *t1*, then Homer is bent, period. And if Homer is straight at *t2*, then Homer is straight, period. But now we have our problem. Our principle, together with the supposition that Homer changes, has led us to affirm that Homer is both bent and straight, period—which is impossible.

The idea that what occurs at a place occurs, *period*, can be formulated more generally and precisely as follows:

9.1   For any x and t, if x is Φ at t then x is Φ.

We discussed a similar principle (5.10) in Chapter 5. (Note that I have not reinstituted the policy of formatting tenseless uses of verbs in all-caps. I assume that by now context should make it clear when that is intended.) When we discussed 5.10, we saw that eternalists have persuasive reasons for accepting it whereas presentists do not. The same is true here, and for similar reasons. Suppose x is Φ at t. Eternalism implies that x exists and t exists; so it seems also to imply that the event x's *being* Φ exists. But if x's *being* Φ exists, x is Φ—to say otherwise makes no sense. So if eternalism is true, 9.1 is true. Similar reasoning will show that other versions of four-dimensionalism imply somewhat modified versions of 9.1; but to simplify our discussion we

shall leave those aside and focus our attention on (static) eternalism as the most viable brand of four-dimensionalism.

Whereas eternalists are committed to 9.1, presentists can reject it. They will sensibly interpret the claim that x is Φ at t as an imprecise way of saying either that x was Φ when t was present or that x will be Φ when t is present (depending on whether t is supposed to be a past or a future time). Interpreted that way, 9.1 is virtually equivalent to 5.10 which, as we saw in Chapter 5, presentists have good reason to reject. So endorsing presentism is one way of solving the problem of temporary intrinsics. Since we have already discussed presentism in some detail in Chapter 5, I shall not say anything further about it here. Instead, I shall focus on three of the most promising four-dimensionalist strategies for solving the problem.

SOLUTION 1: TIME-INDEXED PROPERTIES

In raising the problem, we assumed that change involves having properties like *bentness* and *straightness* only *temporarily*. One way to solve the problem is to deny this assumption and to say instead that change involves *permanently* having properties like *bentness-at-t1* and *straightness-at-t2*. These latter two properties are time-indexed properties. What does the time-indexing *mean*? It signals that the property is had *only in relation to the time*. The difference between saying that Homer is (untensed) *bent-at-t1* and Homer is (untensed) *bent* is like the difference between saying that jogging is pleasant-for-me and saying that it is just plain pleasant. The former says only that its pleasantness applies to me; the latter suggests that it is pleasant for absolutely everyone. (Usually if one says "Jogging is pleasant", the index to the speaker is understood as implied. So to appreciate the force of the example, imagine someone who says "Jogging is pleasant—no, I don't just mean that it is pleasant *for me*; I mean that it is absolutely, objectively pleasant!")

So, according to the proponent of this solution, in our initial setup of the problem we were not as clear as we could have been. Instead of saying simply that Homer is bent at *t1* and straight at *t2*, we should have said the following:

9.2   Homer is *bent-at-t1*.
9.3   Homer is *straight-at-t2*.

Furthermore, 9.1 is not as clear as it could be either. It should be recast as follows:

9.1★   For any x and t, if x is Φ-at-t then x is Φ

But once we have done all of this, we can see two things. First, there is no contradiction in affirming both 9.2 and 9.3. Second, given how we are understanding the time indexing, 9.1★ is false. (9.1★ is analogous to the claim that, for any x and p, if x is pleasant–for–p then x is absolutely pleasant.) One cost to this solution is the apparent loss of unindexed properties like *bentness* and *straightness*. Another potential cost is the fact that change no longer appears to be a "dynamic" phenomenon. Objects change, oddly enough, without ever gaining or losing any of their actual properties.

SOLUTION 2: TIME-INDEXED PROPERTY-HAVING

Just as we might index the properties that objects have, so too we might index the *having*, or *exemplifying*, itself. Suppose that Homer *is* bent. Typically, the italicized term is taken to express the link— property exemplification—between Homer and the property *bentness*; and, typically, this link is understood to involve no reference or relationship to particular times. But one might suppose that this typical understanding is mistaken. Perhaps, for example, there are a great many links between objects and their properties—one for every time. On this way of thinking, it is again never the case that Homer simply *exemplifies* bentness (unless by that we mean that he is now bent or is always bent). Rather, he *exemplifies-at-t_1* bentness, or *exemplifies-at-t_2* bentness, or … etc. There are other ways in which one might try to "index the having"; but we shall focus only on this one.

Indexing property exemplification in this way solves our problem in a way similar to that in which property indexing solved it. It implies that neither the initial setup nor 9.1 were sufficiently clear, and suggests revisions as follows:

9.4   Homer is-*at-t1* bent.
9.5   Homer is-*at-t2* straight.
9.1★★   For any x and t, if x is-at-t Φ then x is Φ

As before, once we have made our revisions we can see both that 9.4 and 9.5 are consistent with one another and that 9.1★★ is false. The costs to this solution, however, are somewhat different. It is less clear that the dynamic nature of change is lost; nor do we lose properties like *bentness* and *straightness*. However, we do end up believing in a great multiplicity of different links between objects and properties where initially we thought that there was only one. Objects do not simply *have* their properties; rather, they are connected with their properties in a more complicated way that involves times. This is puzzling to say the least, and so it might well lead one to look for a better solution.

SOLUTION 3: TEMPORAL PARTS

Consider the statement, "*x* is Φ". So far we have tried solving our problem by indexing the property (Φ) and by indexing the copula ("is"). There is one strategy left to try: indexing the subject ("*x*"). This solution is probably the most popular one. It comes in at least two versions; but for simplicity we will focus on the most widely discussed of the two—the doctrine of temporal parts.

Suppose you have a road that is bumpy for a stretch and then becomes smooth. The road is bumpy; the road is smooth; but nothing is both bumpy and smooth. Contradiction? Hardly. We dissolve this pseudo-problem by noting that the road is bumpy by virtue of having a *part* that is bumpy, and it is smooth by virtue of having a *different part* that is smooth. So likewise, we might say, in the case of change over time. Just as a road "changes" across space by having different parts in different locations with different properties, so too Homer changes over time by having different parts at different times with different properties.

At first this might sound puzzling. Homer is seated at *t1*, standing at *t2*. In what sense does this change involve *different parts*? Here we must distinguish between an object's *spatial parts* and its *temporal parts*. A spatial part of an object at a time is something that exists just at that time and completely occupies only *some* of the region that the object itself occupies. So, for example, your hand right now is a spatial part of you right now: It completely occupies some but not all of the region occupied by you right now. By contrast a temporal part

of an object *x* is something that completely occupies *all* of the spatial regions occupied by *x* during just *some* of the time that *x* exists. So, for example, there is a temporal part of you that includes all of the years between your birth and age 5, another part that covers ages 5–10, another that covers the time you have spent today reading this book, and so on. So too with Homer. He has a seated temporal part at *t1* and a standing temporal part at *t2*. That is the sense in which Homer's change from sitting to standing involves different parts.

As with the other indexing solutions, the doctrine of temporal parts (hereafter DTP) suggests revisions both to the initial setup of our problem and to 9.1. The advocate of DTP will recast the crucial claims as follows:

9.6   Homer-*at-t1* is bent.
9.7   Homer-*at-t2* is straight.
9.1★★★ For any *x* and *t,* if *x*–at–*t* is Φ then *x* is Φ

Homer-*at-t1* and Homer-*at-t2* are both temporal parts of Homer—his *t1* temporal part and his *t2* temporal part, respectively. Since different temporal parts of a thing are different objects, there is no problem with them having different properties. So there is no contradiction in affirming both 9.6 and 9.7. And 9.1★★★ is clearly false. Given DTP, it does not follow from the fact that one of a thing's temporal parts is Φ that the thing itself is (untensed) Φ.

DTP is a view that many philosophers accept for a variety of reasons, not just because it provides a response to the problem of temporary intrinsics. It is, indeed, one of the two main theories about what it means for an object to last over time. The core idea is that objects persist by being *extended* in time in much the same way as they are extended in space. This mode of persistence has been dubbed "perdurance", and so DTP is sometimes referred to in the literature by the label "perdurantism". The contrasting view, *endurantism*, maintains that objects persist by being *wholly present* at every moment at which they exist.

Perdurantism is the theory of persistence that seems to fit most naturally with the ever–popular eternalist theory of time. As we have seen, eternalism implies that time is "extended" in a way that blurs

the alleged differences between time and space; and it is natural, on this picture, to think of a persisting material object as similarly extended in spacetime, and divided—like a loaf of bread—into numerically distinct slices at each moment along its career. Some philosophers even think that eternalism *entails* perdurantism. I disagree, but it would take us too far afield to pursue that issue in detail here.

Perdurantism also solves some of the puzzles that raise the problem of material constitution. For example, consider again the puzzle of Tibbles and Body-minus. The puzzle arises because it *looks* as if Body-minus is a mere part of Tibbles at one time, but is identical with Tibbles at a later time. According to the doctrine of temporal parts, however, Tibbles and Body-minus are both four-dimensional objects, spread out in time just as they are spread out in space. Tibbles fills the region of spacetime occupied by her body over the course of her entire life; Body-minus, however, fills a different region, the one occupied by Tibbles's torso throughout its entire life. So, if DTP is true, there is no pressure to say that Tibbles and Body-minus are ever identical, for it is eternally true, on this picture, that they occupy different regions of spacetime and have different spatiotemporal parts.

That said, however, perdurantism does *not* provide a general solution to the problem of material constitution. Insofar as DTP implies that Tibbles and Body-minus never share *all* of their parts in common, it solves the Body-minus puzzle by implying that Tibbles and Body-minus do not present us with a case of material constitution. But not every constitution puzzle can be solved in this way. For example, consider again the Lumpl/Goliath puzzle. This one was crafted specifically to pose a problem for temporal parts theorists. Lumpl and Goliath coincide for their entire careers. Thus, they share all of their spatial parts at every time at which either exists, and so they share all of the same temporal parts in common. DTP cannot solve this puzzle; thus, in the face of this puzzle, DTP theorists are forced to accept one of the other, more general options listed in Chapter 8.

In addition to having only limited utility as a solution to the problem of material constitution, perdurantism also carries some

counterintuitive consequences. I'll highlight just two of them. First, consider Frank Abagnale, Jr., a security consultant famous for his early exploits as a grifter, check forger, and impostor. (His life provided inspiration for the 2002 film *Catch Me if You Can*.) Most of Abagnale's crimes were committed before he was 30, and most were never prosecuted. Suppose we try to rectify this. We bring him to trial; but he offers the following defense:

> Perdurantism is true—just consider the problem of temporary intrinsics if you need to be convinced. But if perdurantism is true, then, in a deep twist of irony, the individual sitting before you now is yet another impostor! The crimes you wish to try were committed by another man—one we might call the "first thirty years" temporal part of Frank Abagnale, Jr. But *I am not that other man*. More carefully: I am certainly not the "first-thirty-years" temporal part; for I exist now, and he existed only then. Nor am I any temporal part of that man, for the same reason. Perhaps you will insist that I deserve to be on trial because the crimes were committed by the four-dimensionally extended Mr Abagnale, and I am that man. Well, I am not so sure about that. He and I occupy the same space right now. But how do you know that *I*, the man talking to you, am *him* rather than *his current temporal part*? But let me grant you, for the sake of argument, that I *am* the four-dimensional Mr Abagnale. Still, I committed those crimes only by virtue of having that other man—the "first thirty years" man—as a part; and it is not up to *me* what temporal parts I have! I came into this world with him as a part as surely as I came into the world with hands and feet among my parts. My hands and my feet I can cut off, but I can't even cut off my "first thirty years" temporal part. Will you condemn and punish me, then, for something so utterly beyond my control? In the name of justice, then, you must release me.

No court would accept such a speech; but its availability in principle does point to a problem for DTP. The temporal parts of a person—especially the extended ones, as opposed to the instantaneous ones—have all that it takes to *be* persons; so it is hard to see why it would make sense to blame and punish (as we do, if DTP is true) the later temporal parts of a person for the actions of earlier ones. One can insist that in blaming and punishing the parts we thereby blame and punish the person, but that in no way mitigates the fact that in doing so we *also* blame and punish the innocent temporal parts.

The second objection is related to the first. Suppose you reflect on your childhood. You remember your sixteenth birthday, for example. You remember where you were, what you were doing, what the cake tasted like, and so on. Getting a bit more specific, suppose you think at *t1*, "I really enjoyed that cake", and then you think at *t2* "I'm glad that my sister was there". Obviously the transition from the first thought to the second involves a change. Thus, DTP implies that you have different temporal parts at *t1* and *t2*, one of whom thinks "I really enjoyed that cake" and the other of whom thinks "I'm glad that my sister was there". Moreover, it is by virtue of having these two different temporal parts that *you* think those things at the relevant times. But now it looks as if we have *three different thinkers*—you, your *t1* part, and your *t2* part—where in fact we should have only one. This is bizarre. The doctrine of temporal parts implies that for every temporary thought you have, there is some *other thing*—a temporal part of you—that *also* has that thought. Worse, it implies that you have your thoughts derivatively: it is only *because* that temporal part of you has the thought they have that you have the thought that you have. Furthermore, the doctrine implies that some of the things thinking your thoughts have true beliefs and others don't. Your *t1* and *t2* parts never ate your sixteenth-birthday cake (they didn't exist at that time); and your sister is not their sister (they were not born and they have no parents, so they have no siblings; and if they did, they would be your siblings too). Thus, they are mistaken when they think that they enjoyed your cake, and they are confused when they think of your sister as theirs. All of this, I submit, is very odd.

So DTP comes with some costs. However, DTP also comes in different varieties, which avoid some of the costs listed here while taking on other, different ones. For my part, the costs associated with DTP are too high. Endurantism seems a better way to go.

PERSONAL IDENTITY

We turn, finally, to the topic of personal identity. This is a topic on which almost everyone already has at least a few opinions, some grounded in common sense and experience and others perhaps grounded in religious convictions. As I've already suggested, virtually

all of us will think that we are able to survive radical, non-fatal changes in our bodies and minds. We survive growth and development from infancy to adulthood; people survive the loss of limbs and all manner of surgical alterations; our beliefs, emotions, and personalities change over time; and so on. The topic of personal identity comprises a variety of philosophical questions about how such changes are possible and about the conditions under which a person can be said to *survive* from one time to the next.

Some of the questions one might ask on this topic look like nothing more than special instances of the more general questions about change and identity that we have already been discussing. How can one and the same person have different properties at different times? What is the relationship between a person and the mass of particles that constitutes them at any given time? Are persons composed of temporal parts, or not? The puzzles discussed in the earlier sections of this chapter provide a natural starting place for reflection on these sorts of questions. But they provide only a starting place. For persons are not ordinary material objects (if they are material objects at all), and questions about the survival of persons from one moment to the next are in some ways more complicated than questions about the identity of material objects across time.

The questions become most vivid—and most relevant to our ordinary lives—when we consider the way in which human persons are affected by severe brain damage. Those suffering from Alzheimer's disease, for example, commonly experience massive memory loss which is sometimes also accompanied by radical changes in personality. Familiar faces and entire relationships are forgotten; core desires, values, and ambitions may change or be lost altogether; religious beliefs vanish. We find many of these same phenomena present in greater or lesser degrees when people suffer from mental illness, or when they suffer certain kinds of brain trauma. In such cases we find ourselves tempted to say that we are no longer dealing with the same person. We sometimes say that the original person is no longer there. We say these things, too, when a person enters a persistent vegetative state, wherein bodily functions continue to operate (sometimes with assistance) but cognitive activity has ceased. There is no question that there is *some* truth to these claims. But how much truth? Might a person's body survive longer than the person? Can a person survive

apart from their body? These are among the questions raised by the literature on personal identity.

Let us define our topic a bit more clearly. The literature on personal identity has focused primarily on two issues: (i) the nature of persons; and (ii) the necessary and sufficient conditions for *survival*, or *sameness of person* over time. These two issues are obviously related. If what it is to be a person is simply to be an immaterial soul, for example, then a necessary and sufficient condition for the survival of a person from one moment to the next will be the survival of their soul. Similar remarks apply if what it is to be a person is to be an organism, or a brain, or something else. If being a person is just being *an object* of a certain kind *F*, then the necessary and sufficient conditions for *survival* or *sameness of person* over time will just be the necessary and sufficient conditions for the survival of Fs.

The vast majority of thinkers who say that it is in the nature of a person to be an object of a certain kind will go on to say that a person *x* and a person *y* are the same person if, and only if, x = y. In a slogan: survival is identity. Not every theory of persons, however, identifies persons with objects. According to some philosophers, persons are more like software running on the brain, or logical constructs of a certain kind. Others wish to remain noncommittal on the nature of persons but nevertheless have views about the necessary and sufficient conditions for survival. As a general rule, philosophers who do not identify persons with objects tend to think that two individual objects can be the same person even if they are not identical. In a slogan: Survival is not identity. Thus, within the literature on personal identity there are two camps: those who think that survival is identity and those who think that it is not. We shall discuss each view in turn.

We begin with a thought experiment which will help to lend concreteness to our discussion. Daryl Gregory's well-known science fiction story, "Second Person, Present Tense", tells the story of Therese, a teenage girl who suffered complete memory loss as a result of a drug overdose. In rehabilitative therapy, Terry—whose body, at any rate, is the same as Therese's—has recovered some of Therese's memories; but she does not experience them as her own, genuine memories. Rather, Terry experiences them as alien, as if they were implanted pseudo-memories of events in Therese's life. Moreover, Terry's values, core desires, goals, and ambitions are not continuous with Therese's. Accordingly, Terry insists that she is *not the same person*

as Therese. Terry believes this: Therese died with the overdose and Terry was born immediately afterward, sharing only a body with Therese. Therese's parents, however, disagree. Terry is *their child*, and they are hurt by Terry's refusal to recognize herself as Therese and them as her parents. Reflecting on the difference between Terry's and Therese's parents' views about personal identity will help us to sort out some of the most important considerations in deciding between competing answers to the problem of personal identity.

Consider first the perspective of Therese's parents. Why think that Terry is the same person as their daughter? One reason is the simple fact that Terry is the *same (identical) organism* as their daughter. In the story, it is clear that this fact is indeed what moves the parents most in the direction of affirming that Terry is Therese. To take this consideration as decisive is to endorse what is known as the *animalist* criterion of personal identity: To be a human person is to be a human animal; and x is the same human person as y if, and only if, x is the same human animal as y.

The animalist criterion has much going for it. Most importantly, as the following "thinking animal" argument suggests, it seems to have the weight of common sense on its side: Look at yourself in the mirror. What do you see? A human animal. Not only that, but you see a thinking human animal. Do you see any other thinking thing there in the mirror? Surely not. But *you* are thinking. So you are that thinking animal.

This is a persuasive argument, but it is hardly decisive. Some philosophers, for example, believe that it is in the nature or essence of a human person to be or to have a particular *soul*. Some say that persons *are* souls, and that a person's body is just an inessential appendage or tool used by the soul. This sort of view is commonly associated with a variety of ancient Greek philosophers, Pythagoras and Plato among them, and with the 17th-century French philosopher René Descartes (1596–1650). Others say that the relation between body and soul is more intimate, so that even though the soul can exist apart from the body, its natural state is to be embodied and the person whose soul it is cannot exist in a fully functioning state unless their soul is embodied. This latter view is commonly associated with St Thomas Aquinas (1225–1274). On both views, personal identity is tied to the soul. More exactly: x is the same human person as y if, and only if, x is or has the same soul as y. Since a soul can exist apart

from any particular human animal, proponents of this same soul criterion will deny the animalist claim that what one sees in the mirror is, strictly speaking, a thinking human animal. To be sure, it is true in a rather loose way of speaking that human animals think. But *properly* speaking, what thinks is the person, i.e., the soul.

There are other reasons to reject animalism, even apart from belief in immaterial souls. Some do so on the basis of thought experiments like the following. Imagine a procedure whereby your cerebrum (the part that supports thought, not bodily function) is transplanted into a new body, and someone else's cerebrum is then transplanted into your original body. Let's call the organism that has your cerebrum "A", and let's call the other organism "B". Now suppose that, in advance of the procedure, you are told that one of the two resulting organisms will receive your heart's greatest desire and the other of the two will also receive something nice, but nothing like your heart's greatest desire. You are then asked to decide which organism receives which prize. Isn't it *obvious* that you should arrange for A to receive your heart's greatest desire? Clearly so; but it seems likewise clear that A is not the same animal that you were at the beginning of the procedure. (B seems to have the better claim to being the same animal.) Indeed, if one is willing to venture a bit further into science fiction, the case can be strengthened. Just imagine a procedure whereby your entire brain is removed and rigged up so as to control a wholly synthetic body. It seems possible that you survive such a procedure; but at the end of it you would not be an animal. Thus, the animalist criterion again seems false.

The criterion suggested by this thought experiment is a *same-cerebrum* criterion: x is the same person as y if, and only if, x is the same (identical) cerebrum as y. Moreover, one might motivate this criterion not just with thought experiments but with what Hud Hudson calls the "elimination principle":

> Elimination Principle: If x and y are both human person candidates and at most one of x and y is a human person, but y has superfluous parts whereas x doesn't, then x is the better candidate for the office.

Wherever we find an organism, we are faced with a situation wherein we must recognize *a single human person* from (at least) two candidates—the organism and the cerebrum. Both have what it takes to be conscious; but the organism has "superfluous parts", like hands

and toes, none of which it needs in order to be conscious. The cerebrum, on the other hand, is just big enough to do the job. Thus, by the Elimination Principle, we should recognize the cerebrum and not the organism as the human person.

Each of the views just discussed falls into the "survival is identity" camp. Each identifies a human person with some specific object—an organism, a soul, a cerebrum—and each maintains that x is the same person as y only if $x = y$. Moreover, endorsing any one of these views would provide Terry's parents with an argument for the conclusion that Terry is the same person as their daughter. For Terry has the same brain as Therese, she is the same organism as Therese, and (assuming souls are substances that can survive the loss of any and all of their memories) there is no reason to doubt that Terry has the same soul as Therese.

What about Terry's perspective? Taking her perspective does not *require* that she deny that survival is identity. For example, she *could* maintain that given the psychological changes she's undergone, a new soul must have taken up residence in her body. That is an unlikely view, however. Terry's perspective is much more likely to be endorsed by someone who maintains what is known as a *psychological continuity* criterion of personal identity. This sort of criterion is typically endorsed by those who do not identify persons with material or immaterial objects and who deny that survival is identity.

The psychological continuity criterion is most widely associated with the 17th-century British empiricist John Locke (1632–1704), and also with the contemporary philosopher Derek Parfit. One version of the psychological continuity criterion is as follows:

(P1) x is the same person as y if, and only if, x's psychological states display the right sort of continuity with y's psychological states.

Parfit, however, takes sameness of person to come in degrees, which suggests a different formulation:

(P2) x is the same person as y just to the degree that x's psychological states display the right sort of continuity with y's psychological states.

What counts as "the right sort of continuity"? It is hard to say precisely; but any decent answer to that question will have to be consistent with basic intuitions such as these: (a) I am the same person as the 5-year-old boy whose first day of kindergarten I remember, and (b) no matter how much the mental lives of my contemporaries happen to resemble my own mental life, I am not the same person as any of them.

Parfit himself makes a helpful distinction between psychological *continuity* and psychological *connectedness*. A person *x* is psychologically connected to a person *y* if *x* seems to remember events that *y* actually experienced, or *x* sees their actions as carrying out intentions that were in fact *y*'s intentions for their future self, or other mental states of *x* stand in other relevantly similar relations to mental states of *y* or events in *y*'s life. And a person *x* stands in the relation of *psychological continuity* with *y* if, and only if, *x* and *y* are both links in a chain of psychologically connected persons (i.e., *x* is psychologically connected to *a* who is psychologically connected to *b* … who is psychologically connected to *y*). In the normal case, *x* and *y* will be the same person only if *x*'s psychological makeup is related to *y*'s by a series of psychologically connected stages that looks like the development of a single mind over time.

Perhaps the most widely discussed contemporary motivation for the psychological continuity criterion is Parfit's famous double brain transplant scenario. According to Parfit, there is no question that a human person can survive the removal of one hemisphere of their brain. Moreover, he says, technology presents the only obstacle to surviving a hemispherectomy followed by a transplant of the resulting half-brain. So now imagine a case where half of one's brain is removed and transplanted and the other half is *also* removed and transplanted. Let Scarecrow be the person undergoing the procedure; let S1 be the person who results from the transplant of the left half of Scarecrow's brain and let S2 be the person who results from the transplant of the right half. Given that a person can survive the removal and subsequent transplant of either half of their brain alone, it seems that we must admit that Scarecrow survives this procedure. After all, she would have survived if the right half had been destroyed rather than transplanted, and she would have survived if the left half had been destroyed rather than transplanted. Why, then, should the

fact that *neither* is destroyed make it the case that she doesn't survive? But if she *does survive*, then she must survive as S1, as S2, or as both S1 and S2. According to Parfit, it would be wholly arbitrary to suppose that Scarecrow survives as just one of the two. Thus, she must survive as both. But *she cannot be identical to both*. Thus, survival is not identity; and the best way to make sense of our intuitions about survival under the supposition that survival is not identity is to adopt a psychological continuity criterion.

According to Parfit, then, psychological continuity and connectedness, not identity, are what matters in survival. Moreover, Parfit thinks that whether x is the same person as y is a matter of degree; and, indeed, the facts about sameness of person are sensitive to context and interest. So, for example, suppose that Scarecrow undergoes the double-transplant procedure, and then S1 and S2 are genetically modified so as to enable them each to regrow their missing hemisphere over the course of a year. At the end of the year, S1 and S2 each undergo the double-transplant, generating S3–S6, each of whom also regrow missing hemispheres over the course of a year and subsequently undergo double transplants, generating S7–S14. Does Scarecrow survive as S14? Let us stipulate that S14 is psychologically continuous with Scarecrow. Still, she will surely be less so than S2 was. Thus, *if* she is the same person as Scarecrow, she will be so to a lesser degree than S2 was. Moreover, according to Parfit, whether she bears enough continuity with Scarecrow to count as the same person *at all* will depend in large part on our interests, and on the context in which we are asking whether she is the same person. If, for example, we are trying to adjudicate a dispute about property ownership between S7–S14, we might well say that *none* of them counts as Scarecrow. On the other hand, if S14 is the sole survivor among her transplant siblings and we are wondering whether we can award Scarecrow a Nobel prize *by awarding it to her* … well, maybe in that case we would say that S14 is the same person as Scarecrow.

The psychological continuity criterion has enjoyed a fair degree of popularity, but there are serious problems with it. Suppose you are the one to undergo the double transplant, and suppose you are told that just one of the resultant persons will receive your heart's greatest desire whereas the other will not. What would you hope for in that sort of scenario? What would you fear? For my part, I would *fear* that

I would be identical to no one after the transplant, and that the two resulting persons would be people other than me. I would *hope* that my fears are mistaken and that I would be identical to at least one of the resulting persons, preferably the one who receives my heart's greatest desire. These hopes and fears make no sense if our ordinary concept of survival is one according to which survival is not identity. But these hopes and fears *do* make sense. So, our ordinary concept of survival *is* one according to which survival is identity. Contra Parfit, identity *is* what matters in survival.

Might one embrace a psychological continuity criterion of survival while also maintaining (against Parfit) that survival is identity? Such a combination is possible, but implausible. The reason is simple—we can present it in the form of a dilemma. Either mental states have bearers or they don't. That is, either there is a *thinker* for every thought, or there isn't. If there is, and if survival is identity, then a psychological continuity criterion of survival is implausible; if there isn't, then survival cannot be identity.

Suppose that there is a thinker for every thought. Then *persons* are best identified with thinkers. (Persons have thoughts. So if there is a thinker for every thought, what else would we identify the thinkers with if not the persons who have the thoughts?) But it is hard to see how the *identity* of a thinking thing could depend on what it happens to be thinking. But if survival is identity, a psychological continuity criterion of survival implies that the identity of a person *does* depend to some degree on what the person is thinking. (If they have all the wrong thoughts, after all, they will fail to be continuous in the right way with their earlier self.) So if survival is identity, then a psychological continuity criterion is implausible.

On the other hand suppose that there isn't a thinker for every thought. Then at any given moment, there is nothing more to a particular "person" than the various thoughts, sensations, and so on that figure in that person's mental life at that moment. This view was famously endorsed by David Hume. Hume asked what evidence we have for believing in a self, and he concluded that we have none at all. When we introspect, we observe our thoughts and sensations, but we do not observe a self that has the thoughts and sensations. Thus, he concluded, "the identity which we ascribe to the mind of man is only a fictitious one". Indeed, on this picture survival *cannot*

be identity. For it is a truism that persons survive changes in their mental lives; and if survival is identity, this means that one and the same person can exist at multiple times with a different mental life at each time. But on this picture either there are no persons at all (there are just thoughts and sensations, and we pretend that they are persons) or else persons are mere bundles of thoughts and sensations. Either way, persons do not, strictly speaking, exist at multiple times with a different mental life at each time. If there are no persons at all, then obviously no person exists at multiple times. Likewise, if there is nothing more to a person than a bundle of thoughts and sensations, then every time a thought or sensation is added to or subtracted from the bundle, we have a new bundle and therefore a new person. So if persons are bundles of thoughts and sensations, they might exist at multiple times, but not with different mental lives at those times. Persons therefore do not persist through psychological change

The dilemma just offered also provides the makings of an argument against Parfit's views. Start with the second horn of the dilemma— the view that either there is no self, or the self is a mere bundle of thoughts and sensations. Hume accepted this view because of his empiricist scruples. He wanted to avoid believing in anything for which he did not have empirical evidence. But for those of us who do not share such scruples, this theory about the self has one major drawback: it conflicts with the deeply held intuition that *thought requires a subject*. Indeed, it is unclear that Hume himself managed to escape the grip of this intuition. Consider again the quotation above: "the identity which *we* ascribe to the mind of man …" *We who?* Who does the ascribing? On the no-self view, or the mere-bundle view, there is *ascribing* taking place, but nothing doing the ascribing. What sense does that make?

So if Parfit accepts that second horn of the dilemma, then his view faces a serious problem (even apart from its conflict with the intuition that survival is identity). As it happens, Parfit does not go so far as to say that there is *no* self, or that persons are mere bundles of thoughts and sensations; so he is not explicitly committed to the no-self or mere-bundle view. But it is hard to see how he can avoid accepting one or the other. For he does say that the self is not a substance—it is not to be identified with a soul, a brain, a body, or any other material or immaterial substance in the world. But if the

self is not a substance, then what options are left besides the no-self and mere-bundle views? All of the most viable alternatives would seem to have been wiped off the table.

FURTHER READING

Selections from the writings of Parmenides on the topic of change, along with David Lewis's classic statement of the problem of temporary intrinsics, are reprinted in *Arguing about Metaphysics*. On Zeno's paradoxes, see Wesley C. Salmon, *Zeno's Paradoxes* (Indianapolis, IN: Hackett Publishing, 2001). On the topic of supertasks, see Jon Pérez Laraudogoitia, "Supertasks", in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2011, available online here: http://plato.stanford.edu/archives/spr2011/entries/spacetime-supertasks/.

On change and temporal parts, see Roderick Chisholm, "Identity and Temporal Parts", and Mark Heller, "Temporal Parts of Four Dimensional Objects", both reprinted in *Arguing about Metaphysics*. My own "Temporal Parts Unmotivated", *Philosophical Review* 107 (1998): 225–260 surveys several of the most important arguments for belief in temporal parts and provides replies. For book-length treatments (including discussion of the "stage theory" variant on perdurantism), see Theodore Sider's *Four-Dimensionalism* (Oxford: Oxford University Press, 2001) and Katherine Hawley's *How Things Persist* (Oxford: Clarendon Press, 2001).

Jeffrey Brower offers a novel *endurantist* solution to the problem of temporary intrinsics in his "Aristotelian Endurantism: A New Solution to the Problem of Temporary Intrinsics", *Mind* 119, no. 476 (2011): 883–905. This is the second version of the "indexing the subject" solution discussed in the text in connection with "Solution 3". Among the more important defenses of presentist replies to the problem of temporary intrinsics are Trenton Merricks, "Endurance and Indiscernibility" *Journal of Philosophy* 91 (1994): 165–184 and Dean Zimmerman, "Temporary Intrinsics and Presentism", in Dean Zimmerman and Peter van Inwagen (eds), *Metaphysics: The Big Questions* (Cambridge, MA: Blackwell, 1998), 206–220. The time- and copula-indexing responses are discussed in Peter van Inwagen, "Four Dimensional Objects", *Noûs* 24 (1990): 245–256, Sally Haslanger, "Endurance and Temporary Intrinsics", *Analysis* 49 (1989): 119–125, and David Lewis, "Tensing the Copula", *Mind* 111 (2002): 1–14.

On personal identity, the article by Derek Parfit that was cited here—"Personal Identity"—is reprinted in *Arguing about Metaphysics.* So too is Daryl Gregory's "Second Person, Present Tense", and Eric Olson's "An Argument for Animalism", which is the source of the "thinking animal" argument discussed earlier. In

discussing replies to Olson's view, I quoted a principle from Hud Hudson's "I Am Not an Animal", in Peter van Inwagen and Dean W. Zimmerman (eds), *Persons: Human and Divine* (Oxford: Oxford University Press, 2007), 216–234. Two classic, and highly accessible books on personal identity are John Perry's *A Dialogue on Personal Identity and Immortality* (Indianapolis, IN: Hackett Publishing Co., 1978) and Sydney Shoemaker and Richard Swinburne, *Personal Identity* (Oxford: Basil Blackwell, 1984). For a very different way of thinking about personal survival—one which has had a great deal of influence in the literature on human identity and *agency*—see Marya Schechtman, *The Constitution of Selves* (Oxford: Oxford University Press, 1996).

All of these issues intersect with issues about composition, so the readings recommended at the end of Chapter 8 will also be of value to readers interested in the topics covered in the present chapter.

FREEDOM

Whatever will be, will be. Most often, these famous words simply express a carefree attitude toward the future—"Let things unfold as they will; whatever will be, will be!" Sometimes, however, they express an attitude of wistful resignation to whatever fate has ordained—"Nothing I do can make any difference to what happens; the future is set, and there is nothing I can do about it; I can't do anything except sit and wait to see what happens; whatever will be, will be". It is this latter sense of those words that is of interest for purposes of this chapter. Is the future inevitable and unchangeable? Is it true that there is nothing we can do about the future? If so, doesn't this mean that we are not free? These questions point us toward one of two philosophical problems concerning freedom of action that we shall discuss in this chapter: the *fatalism problem*.

We rarely *act* as if we believe that we are fated to act in a particular way. We deliberate, sometimes long and hard, about what we ought to do in various circumstances. In doing so, we manifest our belief that we actually have a choice about what we do, that how things will turn out depends on what we now choose, and that the future is not already settled. Likewise, we take ourselves to be responsible for our actions—as if it is genuinely up to us whether we do the things we in

fact do. If we sincerely believed that our lives are subject to fate, how could we sensibly take ourselves to be responsible for what we do?

Yet we often talk about the future as if there are now facts (i.e., true propositions) about what we and others will be doing. We routinely say things like, "I will be home for another hour; feel free stop by anytime", or "Kristina will babysit Matthias and Penelope this afternoon, so let's go get coffee at one o'clock", or "Tim will leave for the airport in a minute, so you had better quickly say your goodbyes". Moreover, it seems that we can look back on the past and make retrospective inferences about what future-tense claims were true prior to certain dates. Buzz Aldrin took his famous moon walk on July 21, 1969; therefore it seems that the following must have been true at every time prior to that date: "Buzz Aldrin will walk on the moon on July 21, 1969". But if there are future-tense facts like this, then it seems that the future *is* already settled, our current deliberations can make no difference in what will come to pass, and we have always been fated or destined to act exactly as we do.

So there seems to be a tension in the beliefs that many of us have about the "openness" of the future. We think and act as if we have many available alternatives; but we also seem to think that there are facts about what we will do and about how things will go. It is unclear how these two otherwise commonsensical beliefs can both be correct. This apparent tension is what gives rise to the fatalism problem.

The second problem about freedom is related to the first in that it, too, arises out of beliefs many people have that conflict with the commonsense belief that freedom involves having some kind of *control* over our actions. This is the *determination problem*. It arises because many have thought that our acts are determined in various ways, and these forms of determination all seem to preclude our having genuine control over our actions.

Some people believe that the laws of nature are deterministic—that the universe is like a machine that *has to* unfold in one very specific way. Others reject this view about the laws of nature but think that human action is subject to various other kinds of determination. For example, we seem to have no direct control over our beliefs and desires at any given moment. (If you think otherwise, then I have a terrific diet plan for you: just stop desiring unhealthy foods.

I also have a terrific plan for quitting smoking: just stop desiring to smoke. This plan also works for low self-esteem: just start believing that you are a wonderful person and that everyone likes you.) But our beliefs and desires together seem to determine all of our actions. Likewise, we have no control over the laws of physics and chemistry that govern the behavior of our brains which, in turn, determine the behavior of our minds and bodies. All of these determination theses threaten the idea that we have genuine control over our actions. What is worse, rejecting them doesn't seem to help with the problem. For if neither our beliefs and desires nor the laws of nature determine our behavior, what else besides sheer chance could explain why we do what we do? But chance events are no more up to us than the laws of nature. So it seems that the threat to our belief that we control our actions lingers even if we reject all of the above determination theses.

TERMS AND DISTINCTIONS

Let us begin with some of the central "isms" in the free will literature. There are six of them: **determinism**, **indeterminism**, **compatibilism**, **incompatibilism**, **libertarianism**, and **hard determinism**.

Determinism is commonly defined as the thesis that there is, at any given moment, only one physically possible future. We might also put it this way: Let's say that a *world-statement* is a statement describing the complete state of the world at a particular time—i.e., one that fully describes how everything *is* at that time, but omits any truths about how things *were* or *will be*. (A complete state of the world at a time is just like an abstract time, except that the time would include whatever facts there are about how things were and will be. A world-statement, then, is just the proposition corresponding to that world-state.) Let's say that a law-statement is a statement that describes *all* of the laws of nature that hold in the world. Then determinism is the thesis that a world-statement for a time *t* together with a law-statement logically entails *every* world-statement for every time later than *t*.

The other "isms" are easy to characterize. *Indeterminism* is the thesis that determinism is false. *Compatibilism* is the view that freedom is compatible with determinism—i.e., it is possible for us to act freely even if determinism is true. *Incompatibilism* is the view that freedom is incompatible with determinism. *Libertarianism* is the view that

incompatibilism is true and human beings are free. *Hard determinism* is the view that both incompatibilism and determinism are true.

What does it mean to be free, or to act freely? I have no informative definition to offer, but we can home in on our concept of freedom by looking at three other concepts with which it is closely connected. First, the concept of freedom is connected with the idea of control: to be free is to have a certain kind of control over one's acts. Having control over one's acts, in turn, requires being (in some sense) their source. To the extent that we act under coercion, for example, our freedom is undermined. Second, the concept of freedom is connected with the notion of responsibility: our free acts seem just to be those acts for which we can be held responsible. (This is not to say that every free act is morally significant, though. One can be responsible for one's coffee sipping acts, even if taking a sip of coffee is wholly morally neutral.) Third, the concept of freedom is connected with the idea of being able to choose among incompatible courses of action. Where we are unable to act otherwise, we seem not to act freely.

Together, these claims give us a pretty good idea of what we mean when we say that an agent acts freely. Each of them, however, is problematic. The claim that freedom requires control is imprecise, as is the claim that having control involves being the source of one's actions. Does having control require having alternative possibilities for action? If you could not have done otherwise than what you in fact did on some occasion, does it follow that you lacked the right sort of control over your action? Does being the source of your actions mean being subject to *no* outside influences, or just relatively few? If the latter, then which outside influences undermine freedom and which do not? These questions are hard to answer. The claim that our free acts are just those for which we are responsible is not quite so imprecise, but it is controversial. Some philosophers think that freedom and moral responsibility come apart: one can freely perform an act for which one is not morally responsible, or vice versa. Likewise with the claim that freedom requires the ability to do otherwise. There is, as one might guess, a vast literature exploring all of these issues.

Trying to conduct our discussion in a way that remains neutral on the issues just mentioned is unworkable. It would make things too

complicated. Likewise, trying to settle the issues by way of argument is unworkable. It would take a long book rather than just a section of a chapter. So I will simply make some decisions. Following Alicia Finch, I will say that, necessarily, a person S performs an act A freely only if each of the following claims is true:

- *S has or had a choice about whether* S performs A.
- It is or was *up to S* whether S performs A.
- S is or was *able to A and able to refrain from A-ing*.
- It is or was true that S *can A and S can refrain from A-ing*.

(My wording here isn't exactly Finch's, but I am quite sure that the modifications I have made are in the spirit of what she intended.) In the remainder of this chapter, I will use the italicized locutions inter- changeably, taking them to be equivalent to one another. In doing so, I set aside talk about "control" or about being the "source" of one's actions as ways of understanding freedom, and I come down in favor of the claim that freedom requires alternative possibilities for action. I shall also assume that freedom goes hand-in-hand with responsibility—that the acts for which we are responsible are just those with respect to which we are free. Making these assumptions does not substantially affect the content of the discussion that fol- lows; it just prevents the discussion from getting overly complicated. Readers interested in a fuller exploration of these issues, together with considerations for and against the various assumptions I am making here should see Meghan Griffith's *Free Will: The Basics*.

In addition to asking what freedom is, one might also ask what things are properly said to *be* free (or unfree). I have already indicated that I think *actions* are the sorts of things that can be free. But people also tend to talk about the *will* being free, or about *agents* being free, or about *decisions* or *choices* being free. Can all of these things be free? Or just some of them?

Decisions and choices seem to be types of action; so there is no obvious problem with treating them as potential bearers of freedom. We might say that an *agent* is free just when certain actions of theirs are free. Which actions? The answer depends on context. Are prison- ers free? Generally—i.e., in most contexts—we will say no, because there are a great many actions that prisoners are unable to perform because of their confinement. But, of course, even prisoners still

manage to act freely. In prison, one still has a choice about whether to obey the guards or resist them, about whether to participate actively in the riot at the end of the week or try to sit it out, about whether to pick up the telephone handset to talk to one's visitor or just to sit there and gesticulate silently, and so on. In contexts where it is clear that it is these sorts of actions we have most saliently in mind, we might well say that even prisoners are free. Still, though we can make sense of talk about agents being free, our discussion will be more rigorous if we avoid it and focus instead on actions.

What about the will? Not one's *willings*, which are acts just like decisions and choices are acts, but the *will itself*. Is the will free? John Locke (1632–1704) famously argued that the question whether the will is free rests on a confusion. According to Locke, the will is nothing but a *power*—a power to choose. Furthermore, according to Locke, freedom is also a power—a power to act as one chooses. Thus, on his view, to say that a person's will is free is to say that their power to choose has a certain power to act. But this is nonsense. Mere powers to choose cannot themselves have powers to act. So Locke concludes that it is simply confused to ask whether the will is free. Is he right? That depends to some extent on whether he is correct in thinking that freedom and the will are both powers. Freedom does seem to be a power or capacity of some sort. But one might think of the will as something more like a decision-making module within one's mind or brain. Thought of in this way, a person's will would be a *part* of them, and it might make as much sense to talk of the powers of the will as it makes to talk of the powers of one's brain or heart or kidneys. So, in short, it is unclear whether we should follow Locke in saying that it is confused to talk about freedom of the will. Regardless of what we decide, however, it seems that an agent's will is free just to the extent that their actions are free. So, it is ultimately unnecessary to focus on freedom of the will itself, and we shall avoid doing so for the remainder of our discussion.

THE DETERMINATION PROBLEM

The determination problem arises because human freedom seems impossible regardless of whether determinism is true, and regardless of whether our actions are subject to various kinds of local determination.

In this section, I will explain the problem more fully, and then I will examine several of the most well-known strategies for resolving it.

Let us note, first, that throughout this chapter our focus shall be restricted to the actions and freedom of *natural* agents—i.e., human beings and other creatures in the natural world, rather than, say, God or angels. So, unless otherwise indicated, talk about *agents* is to be construed as talk about natural agents. This focus makes sense given that we have defined determinism by reference to the laws of nature, which means that it is a thesis about the natural world.

Now let's start with global determinism. The standard argument for incompatibilism is the Consequence Argument. There are various different versions of this argument, but the following is probably the simplest and most straightforward. Let us assume that, necessarily, if agents exist, it is not the case that agents *always* exist—i.e., there are times before the first agents come into being. (This assumption helps us to avoid some unnecessary complications.) Let us stipulate that times in the "remote past" are times before any agent ever existed. Given all of this, we may argue as follows (adapted from Peter van Inwagen's *An Essay on Free Will*):

10.1   Determinism is true. (Assumption)

10.2   If determinism is true, then no matter what act any agent performs, it is a necessary consequence of the law-statement and the world-statement for a time in the remote past that they perform that act. (From the definition of determinism)

10.3   For any time in the remote past, no agent has or ever had a choice about the truth of the conjunction of the world-statement for that time and the law-statement. (Premise)

10.4   An agent has or had a choice about a necessary consequence of a proposition only if they have or had a choice about whether the proposition itself is true. (Premise)

10.5   Therefore, no agent has or ever had a choice about whether they perform any of their own acts. (From 10.1–10.4)

10.6   An agent acts freely on some occasion only if they have or had a choice about whether they perform the act on that occasion. (From our earlier stipulations about freedom.)

10.7   Therefore: No agent ever acts freely. (From 10.5, 10.6)

If sound, this argument establishes incompatibilism. Determining whether the argument is sound is primarily a matter of determining whether premises 10.3 and 10.4 are true. 10.1 is just an assumption to get the argument going. 10.2 and 10.6 are obviously true given the definitions and stipulations laid out in the first section of this chapter. 10.5 has to be true if 10.1–10.4 are true. So all of those premises are immune to substantive criticism. If one wants to avoid the conclusion of this argument, one has to resist 10.3 or 10.4.

Let us begin by examining 10.4. It is easy to find examples that conform to 10.4. Suppose Fred is a bachelor. It is a necessary consequence of this fact that Fred is both unmarried and male. Sure enough, then, Fred will have a choice about whether Fred is unmarried and male only if he has a choice about whether he is a bachelor. But finding examples that conform to 10.4 isn't at all the same as *proving* 10.4, and it is hard to know how one might prove it. Given this, we should perhaps be a little suspicious of 10.4. But, at the same time, 10.4 has been widely accepted in the literature on free will because it seems quite plausible and it is very hard to find persuasive counterexamples.

10.3 is grounded in the idea that nobody has a choice about the laws of nature or about the occurrence of events in the remote past. Is this true? The answer depends partly on what one thinks the laws of nature are and partly upon what one thinks it means to *have a choice* about something.

Consider first the laws of nature. Laws of nature are regularities in nature; but not just any regularity will be a law. Nothing goes faster than light, and electrons repel one another. These are laws. No human being has set foot on Pluto. That is a regularity about human beings, but it is almost certainly not a law of nature. What is the difference?

Some philosophers say that the difference is some kind of necessity. Humans haven't set foot on Pluto, but they *can*. Nothing has gone faster than light and, furthermore, things *cannot* go faster than light. All electrons (that get sufficiently close to one another) *do* repel one another; furthermore, they *have to*. It is hard to say what this necessity consists in, or what grounds it. It is probably not, for example, what we have been calling *metaphysical necessity*, for (according to most philosophers) the laws of nature themselves are not metaphysically necessary. So it must be some weaker kind of necessity. But never

mind that complication for now. The main thing to notice is that people who think of laws this way will generally agree that the laws are inviolable, except perhaps by divine intervention. Laws, on this view, determine the boundaries of what is physically possible; and nobody has a choice about what is physically possible.

Other philosophers, however, think of the laws as nothing more than regularities that have a certain role to play in our theorizing. The laws are, roughly, regularities whose descriptions have the following property: they are simpler, provide better explanations, and have greater predictive power than descriptions of other regularities, and so they are more suitable than descriptions of other regularities for inclusion in a scientific theory. On this view, a regularity can be a law even if it has exceptions. If, in some far corner of the universe, two electrons attract one another, it might still turn out that the statement, *electrons repel one another,* is, by virtue of its simplicity and explanatory and predictive power, more suitable than any relevant rival for inclusion in scientific theory. So, on this way of thinking about laws, there is no obstacle to saying that one has a choice about the truth of a proposition fully describing the laws of nature. As David Lewis argues in his paper, "Are We Free to Break the Laws", this does not mean that we can *cause* a law of nature to be broken. The laws are not optional in that sense. You can't set out to work miracles and expect to succeed. But, as we have just noted, the laws can have exceptions; and if they had had the right sorts of exceptions in the chain of events leading up to some action of yours, you might have acted differently. So you *could* have acted differently; and had you acted differently, a law would in fact have been broken. You would not have caused the law to be broken, but it would have been broken nonetheless.

One might say the same thing about propositions describing the remote past. The initial configuration of matter could have been different. Or perhaps divine intervention at some later point could have affected the course of events without *changing* the laws of nature. (The laws would have been violated, of course; but, on some views about laws of nature, the truth of a proposition fully describing the laws would remain unaffected.) Had either of those two things occurred, the laws would remain the same, but some of the world-statements

about the remote past that are in fact true would have been false. Had this been the case, you might have acted differently. Thus, you could have acted differently. You would not have *caused* the past to be different, but it would have been different nonetheless.

But does any of this show that one might *have a choice* about the past or the laws of nature? This is where what it means to have a choice about something becomes really crucial. Here is one way of understanding what it means to have a choice about something:

> Choice 1: S has or had a choice about whether *p* is true if, and only if, S has or had at least two options such that (i) it is possible that S choose either of the two options, (ii) if S were to choose one of the options *p* would be true, and (iii) if S were to choose the other option, *p* would be false.

On this way of understanding "have a choice", it seems that the views just described do show how one might have a choice about the occurrence of events in the remote past or about the truth of a proposition describing the laws of nature. For example, suppose you ran a marathon yesterday, and suppose further that determinism is true. Let *p* be the conjunction of a law-statement and a (true) world-statement about a time in the remote past. Now, a compatibilist might say that, prior to running the marathon, you had two options—begin the race or refrain from beginning the race. It was possible that you choose either of the two, so condition (i) is satisfied. Furthermore, we know that if you were to choose to begin the race, *p* would have been true. We know this because you *did* choose the first option, and *p* is true. So condition (ii) is satisfied. Moreover, assuming (as we are permitted to, in setting up our example) that this is an ordinary case of beginning a race, we know that if you were to choose to refrain from beginning the race, *p* would have been false. We know this because, whether determinism is true or not, the laws of nature don't normally allow for someone to begin a race that they have chosen to refrain from beginning. So condition (iii) is satisfied. So, on this interpretation of what it means to have a choice about something, you had a choice about whether to begin the race.

Here is another way of understanding what it means to have a choice about something:

> Choice 2: S has or had a choice about whether *p* is true if, and only if, S is or was able to cause *p* to be true and S is or was able to cause *p* to be false.

On this account of having a choice about something, it seems that the views described above do *not* show how one might have a choice about the remote past or the laws. For I have not described *any* view in the preceding paragraph that tries to explain how we might *cause* either the law-statement or a true world-statement about the remote past to be false. I do not myself think that either Choice 1 or choice 2 provides an adequate account of what it means to say that someone has a choice about the truth of a proposition; so I do not think that our task right now is to pick one over the other. My point is simply that by moving in the direction of an account like Choice 1, one can open the door to compatibilism, whereas by moving in the direction of an account like choice 2, one can hope to keep that door closed.

If you are persuaded by the Consequence Argument, it may be tempting to think that rejecting determinism is necessary *and sufficient* for preserving freedom. But this is not true. There are three further threats that need to be considered. One is a threat from indeterminism itself. Another is a threat from the fact that our acts of will seem to be subject to *local determination*. The third is a threat from the fact that the very idea of *control of the will* seems to be incoherent. I will discuss each in turn.

In "The Mystery of Metaphysical Freedom", Peter van Inwagen offers the following argument (known as the "Replay Argument") for the conclusion that indeterminism is inconsistent with freedom (and, hence, that it is impossible for libertarianism to be true):

> If the laws are indeterministic, then more than one future is indeed consistent with those laws and the actual past and present—but how can anyone have any choice about which of these futures becomes actual? Isn't it just a matter of chance which becomes actual? If God were to "return" an indeterministic world to precisely its state at some time in the past, and then let the world go forward again, things might indeed

happen differently the "second" time. But then, if the world is indeterministic, isn't it just a matter of chance how things *did* happen in the one, actual course of events? And if what we do is just a matter of chance— well, who would want to call that freedom?

(p. 370)

The basic argument here is as follows. If our acts are not fully determined by the causal past and the laws, then they occur partly by chance. If they occur partly by chance, then they are not under our control. If they are not under our control, then they are not free. So, if our acts are not fully determined by the causal past and the laws, then they are not free.

Van Inwagen is perhaps the best known defender of libertarianism. But the Replay Argument, which gives expression to an intuition that many compatibilists before him have cited in favor of their position, moves him in the direction of saying that free will is a mystery. His position seems to be that (i) freedom is *clearly* inconsistent with determinism, but (ii) it is also very hard to see how freedom could be consistent with indeterminism. If this is correct, it looks as if we should conclude that freedom is impossible. So the threat posed by the Replay Argument in conjunction with the Consequence Argument looks rather serious. But van Inwagen has the conviction that we are morally responsible and that moral responsibility requires freedom. Thus, he continues to affirm that we are free but declares freedom to be a mystery. So at the very least the Replay Argument threatens the claim that freedom is intelligible, and at worst it contributes to an overall case for the conclusion that freedom is impossible.

The second threat is the Local Determination Problem. (*Local* determination contrasts with *global* determination. Determinism is a global determination thesis, saying that *everything in the universe* is subject to a certain kind of determination. When we talk about *local* determination, we are talking about the determination of specific phenomena within the universe by relevantly nearby causal factors.) We can deal with this problem relatively quickly because the issues are similar to those at work in the Consequence Argument, and so the relevant responses will be very similar to the responses we

considered in connection with that argument. For the sake of brevity and simplicity, I shall only present an informal version of the local determination argument.

Consider an arbitrary agent, S. Even if determinism is false, S's mental life, at least, is wholly determined at any given time by the physical events that are occurring in their brain. That is, *given* what is taking place in their brain at any given time, they *have* to have the various thoughts, sensations, acts of will, and so on that they have at that time. Moreover, nobody has or ever had a choice about the occurrence of the particular physical events in their brain that determine their own acts of will. We can, of course, influence the events in our brain by, say, looking around, turning on music, drinking caffeine or alcohol, and so on. But we have no idea which specific events correlate with our specific acts of will. So we cannot bring about or prevent particular acts of will by bringing about or preventing particular physical events in our brains. For this reason, then, we have no choice about the occurrence of the physical events that determine our acts of will. But an agent has or had a choice about the necessary consequences of an event only if they have or had a choice about the occurrence of the event itself. So S neither has nor ever had a choice about the particular acts of will that they perform. But if an agent neither has nor ever had a choice about their own acts of will, then they neither have nor ever had a choice about *any* of their acts and therefore they do not act freely. Thus, S does not act freely.

The third and final threat to freedom—the regress argument for the conclusion that it is incoherent to think that one has a choice about one's acts of will—arises when we consider what it would mean to have a choice about one's own acts of will. So far as I can see, you have a choice about your acts of will only if your acts of will are *themselves* the products of your own free acts of will. In fact, if one of your acts of will is not the product of your own free acts of will, there seem to be only two other options for how it might have come into being. It might have occurred spontaneously—i.e., as a matter of sheer chance. In that case it was clearly not under your control. Or it might have been caused by some prior event that is neither identical to nor the product of one of your acts of will. In that case too it is clearly not under your control. So, again, your acts

of will are under your control only if they are the products of your own free acts of will.

But now we must confront the regress argument. Again, let S be an arbitrary agent. The following argument reduces to absurdity the assumption that S has a choice about at least one of their own acts of will:

10.8   S has or had a choice about at least one of their acts of will, A. (Assume for *reductio*)

10.9   If S has or had a choice about an act of will, then that act of will is preceded by a distinct free act of will. (Condition on having a choice about one's free act of will)

10.10   Therefore: A is preceded by a free act of will, B (≠ A). (From 10.8, 10.9)

10.11   An act of the will is free only if one has or had a choice about whether it occurs. (Earlier stipulations about freedom.)

10.12   S has or had a choice about whether B occurs. (From 10.10, 10.11)

10.13   B is preceded by a free act of will, C. (Condition on having a choice about one's free act of will)

10.14   If premises 10.8–10.13 are true and are justified in the ways described, then S has or had a choice about A only if A is preceded by infinitely many acts of will. (Premise)

10.15   No human agent performs infinitely many acts. (Premise)

10.16   Therefore: S neither has nor had a choice about A.
        ★★Contradiction

If the argument is sound, then S neither has nor had a choice about any of their own acts of will. If that is right, then it is easy to derive the conclusion that nobody is free:

10.17   S neither has nor had a choice about any of their acts of will. (From 10.16)

10.18   If S neither has nor had a choice about any of their acts of will, then S neither has nor had a choice about any of their acts. (Premise)

10.19   Therefore: S neither has nor had a choice about any of their acts. (From 10.17, 10.18)

10.20 Therefore: S does not act freely. (From 10.19 and our earlier stipulations about freedom)

One might try to reject 10.14 or 10.15, but I cannot imagine either strategy being at all successful. As I see it, there are only two promising replies. One is to reject 10.18, a reply that would also address the Local Determination problem. The other is to reject the argument for the conclusion that an agent has a choice about their acts of will only if those acts are the products of the agent's own free choices. Doing so would leave us with no reason to accept 10.9. As it happens, the two responses are connected: both seem to require belief in something like **agent causation**; and accepting agent causation provides unified way of responding to all three of the threats to free will that I have just presented.

According to the agent causation theorist, it is a mistake to think that our acts are explained either by chance or by causal relations to prior events. Instead, says the friend of agent causation, our acts are explained simply by *us*: agents cause their most basic acts in a direct, immediate way, not by way of choices, beliefs, or desires, and not as a matter of pure chance. If this view is true, the Replay Argument is defanged: the supposition that indeterminism is true does not at all imply that our acts are the product of pure chance. Likewise, the Local Determination problem and the Regress Argument are both defanged. For premise 10.18 is important to both arguments, but the agent causation theorist has no reason to accept it. They might just say that *willing* isn't an act; rather, willing is simply *agent causation*. In that case, there won't be any acts of will, and so one will not have a choice about any acts of will, but it will not follow from this that one has no choice about any of one's acts. Believers in agent causation will also reject premise 10.9 in the Regress Argument. For, according to the agent causation theorist, there is no reason to think that acts of will—if there are such things at all—are produced by preceding acts of will. Rather, they would be produced directly by the agent.

Agent causation is a fascinating idea, but critics say that it is ultimately unintelligible. Suppose S agent-causes their decision to stand up. What, exactly, is the difference between their causing that decision and the decision's just happening? We cannot say that S *does* something to cause it. For that "something" would be a prior event.

What one wants to say is that the decision is caused simply by S's *agent causal activity*. But the trouble is that there is not—and, indeed, cannot be—any object or event in the world with which to identify their agent causal activity. In short, their agent causal activity seems to be *nothing at all*.

THE FATALISM PROBLEM

I turn now to the fatalism problem. The fatalism problem arises out of the apparent fact that, if it was ever true that *S* will do *A*, then from that point forward S has been unable to refrain from doing A. So if it was true long before S was born that S will do A at *t*, then it looks as if S never had, nor ever will have, a choice about whether they do A at *t*—in which case, given our earlier stipulations about what it means to be free, it follows that S does not do A freely. Since it seems that *every* act of ours is such that it was true long before we were born that we will do just that act at just the time when we in fact do it, we have here a general threat to freedom.

Let us put the argument a little more rigorously. Suppose the present time is *t*, and let $t^\star$ be some time one thousand years prior to now. Let S again be an arbitrary agent. Now suppose that S stands up at *t*. We'll take that as our first premise and will proceed from there:

10.21 S stands up at *t*. (Premise)
10.22 If an event *e* happens at *t*, then it was true at every time prior to *t* that *e* will happen at *t*. (Premise)
10.23 Therefore: It was true at $t^\star$ that S will stand up at *t*. (From 10.21, 10.22)
10.24 It is a necessary consequence of 10.23 that S stands up at *t*.
10.25 S never had and never will have a choice about whether 10.23 is true. (Premise)
10.26 An agent has or had a choice about a necessary consequence of a proposition only if they have or had a choice about whether the proposition itself is true. (Premise)
10.27 Therefore: S never had and never will have a choice about whether they stand at *t*. (From 10.23–10.26)

10.28 Therefore: S does not stand at *t* freely. (From 10.27 and our stipulations about freedom)

Obviously "stand up" could be replaced by a name for *any* act that an agent might perform and the argument would reach the same conclusion. So the conclusion generalizes: If the argument is sound, then nobody ever does anything freely.

Is the argument sound? By my lights, the two most natural and promising places to resist this argument are at premise 10.22 and premise 10.25, so I will focus on them in the discussion that follows.

If we reject 10.22, then we can say that, although the event of S's standing up occurs at *t*, the proposition that *S will stand up at t* was never true at any earlier time. Saying this defuses the argument. Those who wish to defend this sort of response generally take one of two routes. Either they say that (with a few exceptions) propositions about what creatures will freely do *have no truth value*, or they say that such propositions have a truth value, but all of them are false. The difficulty with the first option is that it seems to require the rejection of classical logic. One principle of classical logic is the law of excluded middle, which says that, for any proposition *p*, the following is true: *p or not-p*. Where *p* is the proposition that *S will stand up at t*, excluded middle implies that the following proposition is *always* true: *either S will stand up at t or it is not the case that S will stand up at t*. Furthermore, in classical logic, if this disjunction has always been true, then at least one of the disjuncts has always been true too—which is just what proponents of the first response want to deny. Thus, the first response requires rejection of some feature of classical logic. It is, by the way, because of its reliance on these features of classical logic that the fatalist argument we have so far been discussing is commonly called an argument for *logical fatalism*.

Can one sensibly reject classical logic? That is a thorny question. On the one hand, the correct principles of logic are not like principles of etiquette, varying from culture to culture. They are, instead, necessary truths about what follows from what. On the other hand, one *can* question the logical principles that we take to be true, and both the term "necessary" and the term "follows from" can be defined in different ways. So the system known as "classical logic" is not the only logical system from which one can choose. In the end,

the question whether one can sensibly reject it comes down to the question whether alternative systems of logic offer enough theoretical advantages to warrant embracing one of them instead of continuing with classical logic. My own view is that they do not; but there is no way to defend that view in the space available here.

The other way in which one might reject 10.22 is, as noted above, to say that propositions about what creatures will freely do are all false. This view was discussed in the 1960s by A. N. Prior, who attributed it to the early 20th-century pragmatist, C. S. Peirce; but it seems to have been largely ignored until the past decade. Recent defenders include Christopher Hughes and Amy Seymour. The view depends, in part, on making a sharp distinction between the claim that *e will not occur* and the claim that *it is not the case that e will occur*. To say the former is to say that it is now fixed or settled that *e* won't happen, whereas saying the latter is consistent with it's not now being fixed or settled one way or the other whether *e* will happen. Thus, with respect to S and their standing, the proponent of this view maintains that, prior to *t*, both of the following claims are true: It is not the case that S *will* stand up at *t;* and it is not the case that S *will not* stand up at *t*. In other words, at every time prior to *t*, the future includes *neither* S standing nor S refraining from standing at *t*. Obviously this view preserves the law of excluded middle, so that is an advantage; and it fits well with more general intuitions about the openness of the future.

Are there disadvantages to this second strategy? Obviously the view is inconsistent with eternalism. But presentists who endorse this view—like Amy Seymour, for example—will not be at all concerned by this. More troubling for some, this view is inconsistent with the view that there exists a divine being who knows what every free creature will freely do at every time in the future. For, after all, knowing what every free creature *will* do at every future time requires that, for every future time *t* and every free creature *S*, there are truths of the form *S* will do A at *t*. Also troubling is the fact that this view seems to be inconsistent with the possibility of time travel. If a time traveler from the future shows up on your driveway today, it seems that you can infer from this that it is true now that someday in the future they will make a time travel journey. Moreover, if you ask that person what people around them were doing right before they stepped into the time machine, it seems that you can infer from

this all manner of additional future tense truths: that, for example, the time traveler's little brother will be playing video games at just about the time the time traveler steps into the time machine, and so on. But wouldn't it be strange if these were the *only* fixed truths about the future? It seems much more sensible to suppose that, if time travel is possible, 10.22 is true. There are other, more technical difficulties with this proposal as well; but, as with the problems besetting the strategy that involves rejecting excluded middle, exploring those difficulties in detail would take us too far afield.

What about 10.25—the premise that S never had, nor will have, a choice about the prior truth of *S will stand up at t*? Those who reject this premise typically start by making a distinction between propositions that *depend on the future* and propositions that do not. Propositions like *S will stand up at t* fall into the category of propositions that *do* depend on the future; propositions like *S is standing now* do not fall into that category. They then make a distinction between "hard facts" and "soft facts" about the past. The soft facts about the past are ones that depend to some extent upon what is taking place now or in the future. So, for example, between the times $t^\star$ and t the fact that *it was true at $t^\star$ that S will stand up at t* would count as a soft fact, because it would depend on what takes place at *t*. Hard facts, on the other hand, are ones that do not display this sort of dependence. The fact that Brutus stabbed Caesar is, plausibly, a hard fact about the past, since it displays the right sort of independence from things taking place now. (I say "plausibly", though, because it is notoriously difficult to nail down the hard-fact/soft-fact distinction precisely enough to prevent counterexamples.)

But how can the truth of a proposition depend on the future? How could an agent have a choice about *any* fact about the past? The easiest way to answer these questions is to assume eternalism and then observe that the predicate "true at *t*" is (given eternalism) much like the predicate "true in Indiana". One might *say* things like:

(1)   "An earthquake is happening out west" is true in Indiana.

But what one would mean by saying such a thing is that it is just plain true that an earthquake is happening, and that, from the point

of view of Indiana, the occurrence of the earthquake is to the west. Likewise, then, with a claim like *it was true at $t^\star$ that S will stand up at t*: to say this is just to say that, when $t^\star$ was present, "S stands at *t*" was just plain true, and the occurrence of that event was in the future from the point of view of $t^\star$. Once this is clear, the fatalist argument is defanged. For now there is very good reason for thinking that S has a choice about the truth of 10.23—namely, the fact that, at *t,* they seemingly have a choice about the truth of 10.21. In fact, to assume at the beginning of the argument that they don't have a choice about the truth of 10.21 is simply to beg the question in favor of the fatalist's conclusion.

In making this response, I assumed eternalism. Is the response also available to a presentist? I do not think that it is. The problem is simply that there is no way of explaining how *S will stand up at t* could possibly depend on the future at $t^\star$ when (as the presentist believes) future objects and events do not exist. In other words, a proposition cannot depend on the future if there is no future. Thus, despite the problems associated with these strategies, I think that the most promising route for presentists to take in responding to the logical fatalist is to try to reject 10.22.

In closing this section, I should like to note that, although we have focused here on what is commonly referred to as *logical fatalism*, there is a related problem that goes by the label *theological fatalism*. According to the theological fatalist, it is not merely the past *truth* of claims like *S will stand at t* that threatens our freedom, but rather God's foreknowledge of such truths. The strategies for responding to the logical fatalist carry over as strategies for responding to the theological fatalist: one might argue, for example, that there simply aren't truths like *S will stand at t*, since propositions like that either are false or have no truth value. One might also argue that the fact that God infallibly believed at $t^\star$ that S will stand at *t* is as much a soft fact about the past as the fact that it was true at $t^\star$ that S will stand at *t*. Each of these strategies, however, poses new problems in the theological context. For example, the first strategy forces revisions in our concept of omniscience; the second has been seen as requiring some revision in our understanding of the hard-fact/soft-fact distinction (since facts about what God *believed* in the past seem, intuitively, like they should be hard facts). Moreover, there are other strategies for

responding to theological fatalism that don't solve the logical fatal-ist argument. For example, denying the existence of God solves the former problem, but not the latter.

FURTHER READING

In the text, I cited or quoted from two important works by Peter van Inwagen: *An Essay on Free Will* (Oxford: Oxford University Press, 1983) and "The Mystery of Metaphysical Freedom", reprinted in *Arguing about Metaphysics*. I also indicated that I was following Alicia Finch in my choice of terminology. The article of hers on which I was relying is "Against Libertarianism", *Philosophical Studies* 166 (2013): 475–493. I also discussed David Lewis's "Are We Free to Break the Laws?", *Theoria* 47 (1981): 113–121.

The Local Determination Argument I adapted from the chapter on free will in John Searle's *Minds, Brains, and Science*. In connection with that argument, I mentioned the idea of "top down causation". For a defense of that view, see Trenton Merricks, *Objects and Persons* (Oxford: Clarendon Press, 2001).

In connection with fatalism, I cited A. N. Prior, *Past, Present, and Future* (Oxford: Clarendon Press, 1967), Christopher Hughes, "Openness, Privilege, and Omniscience", *European Journal for Philosophy of Religion* 4 (2012): 35–64, and Amy Seymour, "Presentism, Propositions, and Persons: A Systematic Case for All-Falsism" (PhD Thesis, University of Notre Dame, 2015). I also recommend Richard Taylor's, "The Story of Osmo", and the second chapter of Peter van Inwagen's *Essay on Free Will*, both reprinted in *Arguing about Metaphysics.*

Beyond these texts, I think that the best places to start for readers interested in the topics discussed in this chapter are Meghan Griffith's *Free Will: The Basics* (New York: Routledge, 2013) and Robert Kane, *The Oxford Handbook of Free Will* (Oxford: Oxford University Press, 2002). On classical logic and various alternative logics, readers should see *Logic: The Basics*, 2nd ed., by Jc Beall and Shay Allen Logan (New York: Routledge, 2017).

SOCIAL METAPHYSICS—GENDER

Beginning with the last half of Chapter 9, the focus of this book has shifted away from what are commonly thought of as the most abstract, general features of the world (existence and nonexistence, possibility and necessity, time, substance, composition, and identity) and toward some of the most important aspects of human persons—personal identity in Chapter 9, and freedom in Chapter 10. In the present chapter, we take up the topic of gender, as well as an important related topic—the "construction" of objects and kinds that somehow depend on the thoughts and activities of human beings.

Both topics under discussion in this chapter lie in the field of **social metaphysics**, and they are two among several such topics that have lately been receiving a lot of attention from metaphysicians working in the tradition of analytic philosophy. Social metaphysics is concerned primarily with metaphysical questions about the objects, structures, and categories that are central to human social life: gender, race, sex, disability, socioeconomic class, sexual orientation, the nature of groups, collective belief and action, corporate personhood, and many others. Although gender is the main focus of this chapter, I include references to important literature on race, disability, and sexual orientation in the "further reading" section at the end of this chapter.

Topics in social metaphysics were long neglected by analytic metaphysicians, partly because they are not obviously concerned with "fundamental" entities and attributes (recall from Chapter 1 that some want to characterize metaphysics as exclusively concerned with what is fundamental), and partly because they are concerned with entities whose nature and existence seem to depend directly or indirectly on human mental states and social conventions. The idea that some things are mind dependent in this way has seemed puzzling to many philosophers; and, as I will explain a bit more fully later on in this chapter, it has also been seen as a reason to think the things in question don't "really" exist. But I think that, in fact, it is not hard to see how social kinds, at any rate, might be "socially constructed", and therefore at least indirectly mind dependent; and I think that it is likewise easy to see that their being socially constructed does not necessarily imply that they are somehow unreal. To explain all of this, I will devote a portion of this chapter to discussing mind dependence in general, along with the more specific idea of social construction.

The chapter will proceed as follows. I will start by explaining some distinctions that will play a significant role in the discussion that follows: the sex/gender distinction, and the distinction between what I will call "constructed" kinds and "non-constructed" kinds. I will then explain some different ways in which things might count as mind dependent, and I will discuss some different views about the extent to which familiar things might be mind dependent. Having gone through all of this, we will finally be in a position to explore some of the main theories about the metaphysics of gender that that have been discussed in the literature in contemporary analytic philosophy.

SEX, GENDER, AND TYPES OF KINDS

Simone de Beauvoir famously wrote that "One is not born, but rather becomes, a woman". In saying this, she did not mean that people become women in the way that they become adults, or auto mechanics, or asthmatics. Becoming a woman on her view is not a physiological process, nor is it a straightforward matter of deliberate choice and training (though people do indeed receive and choose whether to embrace training in how to be a woman), nor is it a causal

effect of environmental factors interacting with particular features of one's body. People are made into women, on her view, by social norms and widely held beliefs and attitudes that consider certain traits, preferences, behaviors, and social roles to "naturally" go along with being female. Were it not for those social norms, beliefs, and attitudes, *there would be no women*.

De Beauvoir's view is common nowadays, but controversial. Many people think that one's gender (girl, boy, man, woman, non-binary, etc.) is a property that one is born with, and that it is either identical to one's sex (female, male, intersex) or is determined primarily by the properties that determine one's sex, including chromosomes, reproductive anatomy, and secondary sexual characteristics. Others think that sex is a biological given but, similar to de Beauvoir, think that gender depends on social norms, private or collectively held beliefs and attitudes, political structures, or other social phenomena. This idea is sometimes expressed by the slogan, "gender is the social significance of sex". Still others agree that gender depends on social phenomena but disagree with the view (controversially attributed to de Beauvoir) that sex is a biological given. When a doctor or other relevant authority declares that a baby is male or female, or sees to it that one or the other sex is written on an official document like a birth certificate, the declaration depends to some extent on definitions that make reference to anatomical, hormonal, and chromosomal features. As Riki Wilchins notes,

> It turns out birth sex is like a menu. If your organ is less than three eighths of an inch long, it's a clitoris and you're a baby girl. If it's longer than an inch, it's a penis and you're a baby boy. ... But if it's in between, you're [intersex]: The organ is an enlarged clit, and it gets cut off. The pediatrician will apologetically explain to your parents that you were born genitally "deformed," but ... they can make you into a "normal little girl."

Criteria like these are not written into the fabric of the universe; they depend on human choices (which, in turn, depend on beliefs, attitudes, and so on). Nor do they always yield precise or uniform verdicts. One can count as male by anatomical criteria but female by chromosomal criteria, for example, as in some cases of congenital adrenal hyperplasia. Some think that all of this means that sex, like gender, is partially or completely determined by social conventions.

With respect to both sex and gender, then, there is controversy over whether they are **constructed kinds**, that is, kinds whose members are mind dependent, or not. Likewise, we might wonder if a variety of other socially significant kinds—race, disability, sexual orientation, and the like—are socially constructed. There is also controversy over whether they are, or correspond to, natures, or essences. Finally, there is controversy over whether gender kinds, and constructed kinds generally, correspond to real, metaphysically important attributes of things, or whether they are just made-up classifications (perhaps like *nerd*, or *side dish*) that do not correlate with any genuine attribute, or objective or explanatorily useful resemblance among their members. As I will explain more fully in the next section, mind dependence is often taken as a sign of metaphysical unimportance or unreality; and so there is a serious question whether, if genders are constructed kinds, they ought to be accorded a kind of second-class (or worse) status in our metaphysical accounting of the world. Our exploration of mind dependence in the next section will shed some light on this issue.

MIND DEPENDENCE

To say that constructed kinds correlate with genuine attributes of things is to say that human mental activity somehow contributes to making things the way that they are. To say that they correlate with *natures* is to say that human mental activity not only contributes to what things are like, but it helps to determine their very nature and persistence conditions, and even helps to determine the boundaries of what is possible for things in the world. But what kind of mental activity, and determination by mental activity, are we talking about here? What would a mind dependent object be, and what would mind dependence amount to?

Perhaps the most obvious examples of mind dependent things are things that exist *in* minds—thoughts, beliefs, and other mental states. Likewise, anything that is built or otherwise causally brought into existence according to a human design, or as a result of human plans and intentions, depends for its existence in obvious ways on minds. But these are not the philosophically interesting or controversial kinds of mind dependence, so let us set them aside in the discussion

that follows. When philosophers say that something is mind dependent, usually they mean one of two things.

First they might mean that it *directly depends* on human mental states, in that the beliefs, attitudes, social conventions, or other mental states that human beings have individually or collectively adopted partly *make it the case* that the thing exists as the sort of thing that it is. Consider, for example, things like nations or works of art. A nation is what it is in part because it is *recognized* as such by other people; and at least some artworks—most notably found art and abstract art—count as such because they are believed by certain people to be artworks. A group of people can't found a nation simply by writing a constitution, arming themselves, marking borders, and declaring themselves to be a nation; they also need (among other things) recognition from world leaders and even cartographers. Likewise, you can't turn something into a piece of "found art" simply by declaring it to be a work of art. Something's being a work of art depends in part on its being accepted as such by art brokers, critics, collectors, and others in the art world.

Alternatively, in saying that something is mind dependent, one might mean that it *indirectly depends* on human mental states by virtue of depending on laws, social structures, cultural ideologies and conventions, or other social phenomena that are directly dependent on human minds. In other words, something that depends indirectly on minds is something that is at least partly the causal result of the existence or activities of things that are directly dependent on human mental states. Consider, for example, dollar bills, or chess games, or scotch whisky. A piece of paper counts as a dollar bill not solely on the basis of its physical attributes but partly by virtue of its relationship to certain social institutions (a treasury, certain laws, and so on) that depend on complex patterns of recognition, consent, agreement, and other human mental states for their continued existence. A series of physical movements counts as a game of chess in part because the human beings making the movements *intend* to play according to a certain set of rules (rather than, say, the rules of crazyhouse). A bottle of whisky counts as Scotch not just on the basis of its chemical composition but also in part by virtue of its having been made in Scotland (or select other locations), which, like social institutions, depends for its continued existence on complex patterns of recognition, consent,

and the like. These things, then, are good candidates for being things that depend indirectly on minds.

Despite the fact that some of the most important aspects of human social life involve mind dependent things, mind dependence has long been seen by many philosophers as a sign of unreality, a sign that the thing in question is not really even a candidate for serious metaphysical inquiry. Consider again some of the examples from the previous two paragraphs. Isn't it at least tempting to think that, in the case of a dollar bill, the metaphysically interesting thing is the piece of paper, or perhaps the collection of fundamental particles, and not the socially significant monetary unit? In the case of a nation, isn't it at least tempting to think that it is only individual people that are metaphysically interesting, and the rest of what goes into making a nation is just "a bunch of stuff that people do" or something like that? In light of this, it is unsurprising that, as we saw in Chapter 3, realism about objects of a certain kind is usually (although, I should add here, controversially) characterized as the view that there are such objects *and* that an object's being of that kind doesn't depend on human beliefs, opinions, or concepts. Given this characterization, people who believe that social objects and kinds (e.g., dollar bills, or human genders and races) are mind dependent will count as *anti-realists* about such things; and so it is easy to see why one might think that, if metaphysics is concerned with the structure of *reality*, supposedly *unreal* social kinds and their members aren't good candidates for metaphysical inquiry.

The most extreme version of anti-realism is **idealism**, according to which everything either is a mind or is among the contents of a mind. On this view, all familiar kinds are constructed kinds. Idealism received its most famous defense in the 18th century, by Bishop George Berkeley (1685–1753), but it has very few defenders nowadays. A somewhat more modest version, however, grants that there is *something* in the world that is neither a mind nor dependent upon minds, but nonetheless maintains that all of the kinds with which we human beings are acquainted are constructed kinds. This view is often referred to as **constructivism**, since it is a view according to which human minds play an unexpectedly important role in "constructing" the world. The view has been associated (somewhat controversially) with philosophers like Immanuel Kant (1724–1804),

Hilary Putnam, Richard Rorty, and Nelson Goodman. I will not discuss constructivism in great detail here, but I do want to briefly explain it, since doing so will prime us for thinking about how genders might be constructed kinds that nonetheless correspond to real, metaphysically significant attributes of things.

The basic idea is something like this. Just as a piece of paper counts as a *dollar bill* only if it stands in the right relationships to certain human beliefs, concepts, and other mental states, so too a certain mass of matter counts as a *tree*, a *horse*, a *computer*, or a *ship* or whatever only if it stands in the right relationships to certain human beliefs, concepts, and other mental states; and the very concepts that determine *what* a thing is also help to determine its modal properties. A dollar bill could not have been made out of different paper because the concepts, acts of recognition, and other mental states that partly determine what it is to be a dollar bill do not allow for different kinds of paper to constitute dollar bills. So likewise, according to the constructivist, whether a mass of matter makes up a ship that *can* survive complete replacement of its parts or a ship that can't (to revisit the Ship of Theseus example from Chapter 8) depends in part on what is built into the concepts and other mental states that determine what it is that the mass of matter makes up at any given time. Historians may look at the mass that makes up the Ship of Theseus and conceive of it as something that can't survive complete part replacement; sailors, by contrast, might look at the same mass and conceive of it as something that can survive complete part replacement. And, depending on the brand of constructivism one endorses, it might turn out that there is no "objective truth" about what kind of thing the Ship of Theseus is—no truth that is the "same for everyone", so to speak, about what the Ship of Theseus is, or what it can and cannot survive.

Here is an analogy. It is as if the world is a great cloud formation, and we human beings all get to look at it and simply decide what we see in it. When two people look at a cloud formation, it doesn't make sense for them to argue about whether there are "really" three trees, a car, and the face of a woman in the clouds. Nor does it make sense for them to argue about whether the gradual movement of the clouds has objectively created or destroyed any of the objects they see there. For, independently of their thought processes, there is *nothing* objectively there apart from a grand configuration of grey

and white stuff. They see what they see, and each is entitled to their own perspective.

In reflecting on this analogy, it is tempting to think that what is "really" there, objectively speaking, is *the clouds*. It is likewise tempting to think that the trees, cars, faces, and so on that people see in the clouds are *unreal*, for the simple reason that the facts about whether there is a tree, a car, a face, or whatever in the clouds depend on people's beliefs, imaginings, and the like. But I think these temptations should be resisted. For one thing, the perception of *clouds* is just as dependent on our background beliefs and other mental states as the perception of *faces*. We might just as easily have "seen" a mere plurality of water droplets, and drawn no significant conceptual distinction between clouds and other misty formations. Moreover, if there were a cloud formation in which pretty much everyone who looked at it saw a face—e.g., a face resembling that of Lisa Gherardini, the subject of the *Mona Lisa*—it would be very hard to deny that there is "really" a face in the clouds. But it seems that this face is no more or less real, and no more or less dependent on our thoughts and imaginings, than the more inchoate faces and other shapes that people see and disagree about in ordinary cloud formations.

Of course, the situation with familiar material objects like trees, horses, planets, stars, and so on isn't exactly like the cloud formation case. Most importantly, there isn't nearly the same degree of variation in what people see, generally, in the universe as there is in what people ordinarily see in clouds. The case of stars and trees and so on is more like the case of seeing the face of Lisa Gherardini in the clouds: everyone looks at the sky and sees stars; everyone looks at a forest and sees trees; etc. Because of this, there is an important kind of objectivity—again, sameness for everyone—to the judgment that there are stars and trees and so on that is lacking in judgments about what is to be found in a cloud formation. But, according to the constructivist, the difference between interpreting cloud formations and perceiving the reality of trees does not really amount to much more than a difference of degree. For, even if we are built to look out at the world and "divide" it into trees, stars, people, and so on, we might not have been; and we are not the only sensing beings in existence. It is doubtful that bats, for example, or aliens with very different sensory modalities (if there were such things) would divide the world into

the sorts of objects we divide it into. So in that sense the analogy with cloud formations still holds.

Constructivism is a fascinating view. I see its attractions, though I myself cannot ultimately accept it. I am discussing it here, however, because it provides a vivid example of a (widely discussed and defended) view according to which at least some familiar kinds are constructed and, furthermore, constructed kinds are as genuine and metaphysically important as humanly recognizable kinds ever get. It should be clear from the foregoing that constructivism as such gives us no reason to doubt that metaphysical investigation into the nature of persons, or the conditions under which composition occurs, or the persistence of material objects, or the nature of space and time, is somehow pointless or unserious. Interestingly, it does imply that such investigation will need to take greater-than-usual account of whatever diversity we find in people's conceptions of such things since, on constructivism, human concepts are partly determinative of what the world is like. But that is no basis for thinking that metaphysical theorizing about familiar kinds of things is either frivolous or doomed to failure.

Among the most well-known defenders of the view that gender and other social kinds are constructed, not many are thoroughgoing constructivists. I will use the term **social constructionism** to refer to the view that at least some social kinds are constructed; and when I want to refer to views according to which a specific social kind (e.g., gender) is constructed, I will use a term like "social constructionism about gender". Note that, although social constructionism is consistent with both constructivism and the somewhat more limited view that all social kinds are constructed, it does not strictly imply either one.

What social constructionists would say about the kinds they do think are constructed to some extent mirrors the sorts of claims I have attributed to the constructivist, but there are also some important differences. So, for example, those who maintain that gender kinds are constructed typically agree that they correspond to real, metaphysically important attributes; but few of the analytic metaphysicians defending social constructionism about gender think that genders are natures of anything, or that natures correspond, in general, to constructed kinds. Social constructionists about gender also

seem generally committed to the idea that what it is to be of a certain gender, and whether some specific person counts as belonging to that gender, will be at least partly determined by the particular gender concepts operative in some salient group of concept-users. The question, of course, is which groups are salient; and here we find a lot of variation.

If the salient group were *human beings in general*, then gender classifications would be absolutely invariant across time and culture, so that someone who counts as a woman in a certain culture at a particular time would count as a woman in every culture at all times. But few, if any, scholars think that gender is a social construction *and* also think that gender classifications are invariant in this way. Accordingly, every social constructionist about gender that I am aware of seems committed to one or both of the following theses: (i) what it is to be a man, woman, or member of another gender category is not the same for all human beings within those categories; (ii) whether a particular person counts as a man, woman, or member of another gender category is context sensitive—they might be a woman, say, relative to one context but not another. Exactly how these commitments are borne out, however, will obviously depend on the details of the particular theory of gender we are dealing with. At this point, then, it seems best to bring the general discussion of mind dependence and gender to a close and turn to some of the more important specific theories of gender.

GENDER

Let us start by disentangling questions about the meaning of gender terms like "woman" from questions about the metaphysics of the category (or categories) referred to by that word. Although standard English dictionaries offer "adult human female" (or some variation on that) as one of the primary definitions of "woman", it would be naïve at best to think that this settles the question of what it *really is* to be a woman. Dictionaries are usually poor guides to metaphysics, in part because the metaphysical questions one is trying to settle by appealing to the definition often crop up *within* the definition.

Consider an example. The definition of "free" in the *Oxford English Dictionary* that is most relevant to the metaphysics of freedom is as

follows: "Able to act as one wishes, determining one's own action or choice; done or made without compulsion or constraint". As you will know from Chapter 10, however, being able to act as you wish is not quite the same as *determining* your own action or choice. So right away we need to make a decision as to which of these two phrases is primary in the definition—a decision we will *not* make by further appeal to the dictionary, but rather by the usual methods of metaphysics. Moreover, the dictionary does not specify what it takes to determine one's own action or choice, and so (for example) agent–causation theorists, compatibilists, and others are free to inter-pret the definition in accord with their preferred theories. We find much the same situation in the definition of "woman". Dictionary-philosophers of gender treat the dictionary definition as settling the metaphysics of womanhood in favor of the claim that someone is a woman if, and only if, someone is an adult female human. But when (for example) we ask the dictionary itself who counts as *female*, we learn that the adjectival sense of "female" means "[belonging] to the sex which can bear offspring (contrasted with *male*); characteristic of or relating to this sex; cf. FEMININE". This leaves plenty of room for different theorists of gender to interpret "adult female human" in ways friendly to their own particular theory of gender. And, of course, the interpretive room only expands when we consider analo-gous questions about who counts as *adult* and *human* (never mind the further difficulties raised by the fact, noted earlier, that the "biologi-cal" criteria do not always yield uniform or precise verdicts).

I have lingered a bit over dictionary-metaphysics because I think it is a much greater temptation in dealing with the metaphysics of gender than it is in other areas of metaphysics (where we usually can see right away that the dictionary is virtually useless). But I want to turn now to more serious considerations. Perhaps the easiest way to get a handle on the variety of available views and their relationships with one another is to see debates about the metaphysics of gender as focused on two key questions: (a) how many distinct metaphysical categories correspond to each gender word, and (b) what types of categories are we dealing with?

With regard to the first question, an analogy with the litera-ture on freedom will again be helpful. As we saw in Chapter 10, the main competitors in the free will debate are compatibilism and

libertarianism. Strictly speaking, a compatibilist is someone who thinks that compatibilism gives the true story about the one and only type of metaphysical freedom; and a libertarian is someone who thinks that libertarianism gives the true story about the one and only type of metaphysical freedom. But one often finds people talking as if there are *two* types of metaphysical freedom, one whose nature conforms to the compatibilist story, and another whose nature conforms to the libertarian story. On that way of thinking, the salient question is not so much *what is freedom itself like*? but rather *what type of freedom*, *if any, do we have*? Moreover, if there are indeed two types of freedom, then the word "free" is almost certainly ambiguous between them, and we can therefore ask not only which sense is the *dominant* meaning but also which sense is the *better* meaning for theoretical purposes. We can, in other words, ask not only what "free" *does* mean, but what it *ought* to mean.

In the case of gender, most theorists agree that, in the ordinary thought and discourse of culturally dominant groups, the categories of *woman* and *adult female human*, or maybe *adult female person*, are roughly coextensive (but only roughly, due in part to significant ambiguity in each of the three terms). But it is also widely recognized that the culturally dominant meanings of both "woman" and "female" are hardly the *only* meanings of those terms. Talia Mae Bettcher, for example, speaks of "resistant" meanings that are prevalent in at least some **transgender** communities according to which sincerely identifying oneself as a woman is sufficient for being a woman, and having biological markers of femaleness is in no way necessary for being a woman. (Individuals who identify as transgender are usually people who identify themselves as belonging to the gender opposite to the one that is commonly thought to accompany the sex they were assigned at birth. So, for example, a trans woman is someone who was assigned "male" at birth but who identifies as a woman, and a trans man is someone who was assigned "female" at birth but identifies as a man.) For Bettcher, the important philosophical question is not which meaning *is* culturally dominant (this can be settled by empirical inquiry), nor which meaning captures the "true" nature of womanhood (since she seems to think that all of the metaphysical categories corresponding to the multiple meanings of "woman" are genuine categories), but rather which meaning is

better, or which meaning *ought* (ethically, or for political purposes) to come to dominance.

In a roughly similar vein, Sally Haslanger has argued for **ameliorative definitions** of gender and race terms—definitions, that is, that do not aim to capture the dominant meanings of those terms, but are rather crafted so as to serve appropriate moral and political goals, such as the goal of eradicating certain kinds of oppression, doing away with problematic social hierarchies, and the like. At first glance, this might seem like an approach that amounts to just trying to make terms mean what one wants them to mean rather than paying attention to what they "really" mean. But to think of it this way is to misunderstand the project.

To see why, just consider the term "matter" (and the terms in other languages that get translated as "matter"). It is easy to point to regions of space that pretty much everyone will agree are filled with "matter"; but, despite the existence of uncontroversial cases, people have conceptualized matter in very different ways over the course of the past two millennia as philosophers and scientists have theorized about the material content of the world. There are now multiple concepts of matter scattered throughout the history of philosophy and science. Aristotle's concept of matter was different from Isaac Newton's, for example; and both of these are different from the concept(s) in play in contemporary physics. Furthermore, each of these concepts was theoretically useful in its time and respected the generally uncontroversial common judgments about which regions are filled with matter. So each forms the basis for a good candidate definition of "matter". Of course, it's not a live option for contemporary physicists to adopt a Newtonian characterization of matter for their particular theoretical purposes; but it *is* a live option for them (or other theoreticians) to say that (a) Newton's characterization serves our broader theoretical purposes better, and so (b) contemporary physics is offering theories according to which many or all of the regions Newton and others *thought* were filled by matter really aren't. In other words, it is open to them *either* to abandon the Newtonian conception and say that contemporary physics has discovered that matter is not what Newton thought it was, or to retain the Newtonian conception and say that contemporary physics has shown that there is no such thing as matter, or that matter is not distributed throughout the world in

quite the way that we thought it was, or whatever. Obviously my point here isn't to advocate one way or the other on this topic; rather, the point is just that here is a pretty clear case where the question "what do we want this word to mean?" clearly makes sense, and is sensibly answered by thinking about which of the candidate meanings best serves our overall theoretical purposes. To protest that raising this question is just ignoring the "real" meaning of "matter" and trying to make the term mean what we want it to mean is to completely miss this point. So likewise in the case of Haslanger's project.

We come now to the second question: to what kind, or kinds, of category do gender terms refer? Note that this is a question for both multiple-meaning theorists like Bettcher and single-meaning theorists of various sorts. Here we find four main options.

First, gender might be a **real essence** (as contrasted with a merely **nominal essence**). A nominal essence is an abstract idea, grounded in superficial similarities of things, that we use as a basis for classifying them. The categories of *pizza* and *salad*, for example, plausibly correspond to merely nominal essences. There is no deep underlying structure shared by all and only pizzas, or all and only salads—trying to come up with one makes for a fun, if rather nerdy, party game. Instead, things count as pizza or salad simply by virtue of their superficial similarity to some paradigms. By contrast, a real essence is an intrinsic, complex attribute, or underlying structure, shared in common by the members of a kind, which serves to explain their superficial similarities and scientifically interesting features (e.g., behavior, natural development, overall appearance, and so on). As we saw in Chapter 7, Aristotle thought that essences like this were definable in terms of necessary and sufficient conditions; but others—for example, John Locke (to whom we owe the contrast between real and nominal essences)—do not.

Gender essentialism is usually characterized as the view that gender attributes are real essences. Although there are many different ways in which gender essentialism might be developed, and a variety of subtle ways in which the essentialist views of diverse historical figures differ from one another, the following represents a typical and, I think, very familiar essentialist package. The gender attributes *woman* and *man* are biologically determined psychological profiles; thus, they go hand-in-hand with the biological sex traits,

female and *male* (respectively). In the normal and appropriate course of things, being female results in a person developing a distinctively feminine nature. Feminine nature, the true essence of womanhood, is characterized by (among other things) a propensity toward feeling and emotion as contrasted with reason, toward passivity and submission as contrasted with agency and leadership, toward weakness and dependency as contrasted with strength and independence, and toward nurturing and compassion as contrasted with the sorts of "harder" qualities one needs to lead men into battle and accomplish the grimmer necessities of life. Being male, in turn, results, in the normal and appropriate course of things, in possession of the contrasting attributes. On this picture, then, there is a kind of complementarity among the genders. Masculinity and femininity harmonize with one another by virtue of their complementary strengths and weaknesses. Moreover, the differences between men and women are purportedly both genuine and essential, grounded in the biological natures of males and females respectively; and they are held to go a long way toward explaining the personality traits and preference clusters that "we" (white Westerners in the first instance, given the authors of and target audiences most commonly associated with this brand of essentialism) think of as characteristic of the two sexes.

Gender essentialism doesn't *have* to take the patriarchal (man-privileging) form of the account just described; but it commonly has, and this fact is an important part of the explanation why feminist scholars have tended to oppose gender essentialism. But there are also more general reasons for opposing gender essentialism. Perhaps the most important one is methodological. Historically, (mostly male) philosophers have tended to theorize about maleness and femaleness, and masculinity and femininity, in a way that seems mainly to rely on general intuitions about what men and women essentially are and ought to be, and on casual observation of men and women in the theorist's own country, culture, and social class. But such a method, deployed by men in the patriarchal cultures that have dominated throughout the Western philosophical tradition, is virtually guaranteed to be tainted by background beliefs and stereotypes that favor men in various ways, as well as by sampling bias that misleadingly suggests that *common* attributes of men and women living under patriarchy are *natural* attributes of men and women in general. The

result has been seriously harmful. In the work of essentialist philosophers throughout history, we find metaphysical and moral theories that privilege stereotypical attributes of men and embrace negative stereotypes about women (e.g., theories according to which reason is superior to passion, and reason tends to dominate in men whereas passion tends to dominate in women). These theories, in turn, have resulted in a variety of genuinely negative attributes (e.g., lack of self-control) or culturally disvalued attributes (e.g., emotionality, as contrasted with dispassionate rationality) being seen as *natural* traits of women that *justify* their lower social position, when in fact they are not traits of women in general at all, and are at best traits prevalent in a particular population of women in large part *because of* their lower social position and a variety of injustices perpetrated against them. (Readers interested in specific examples and discussion of the harms just described should see the works by Rousseau, Spelman, and Wollstonecraft listed in the Further Reading section at the end of this chapter.)

The main alternative to saying that gender is a real essence is (as we have seen) to say that it is a constructed kind, one whose membership is at least partly determined by people's beliefs about or attitudes toward the things belonging to the kind. Note that in saying that gender is not a real essence (in the sense defined earlier) social constructionists are not committed to the idea that gender is *unreal*; nor are they committed to the idea that gender *has no essence*—i.e., that there is no such thing as "the nature of gender". In fact, those who believe that attributes have essences or natures will regard social constructionism as just one among several theories *about* the nature of gender; and, though some social constructionists might aptly be characterized as anti-realists about gender, plenty will agree that gender is real and will furthermore insist on a characterization of realism about gender that omits affirmation of "mind-independence" as a criterion for realism and that locates them firmly in the realist camp.

Although social constructionists agree that gender is a constructed kind, it is important to bear in mind that there are different types of constructed kinds, as well as different stories about how exactly the relevant "construction" has arisen and is supposed to work. Accordingly, there are different types of social constructionism about gender. Here I will briefly sketch three of the most important

varieties—the second, third, and fourth of the "four main options" on the landscape of theories about gender.

Social position accounts of gender identify gender categories with positions in a social hierarchy. The basic idea is that people occupy different social positions in a society as a result of the interaction between what others believe their sex to be and the various positive and negative attitudes, stereotypes, social norms, and so on that are generally associated with each of the sexes in that society; and having a particular gender is just a matter of occupying one of these different social positions for the reasons just stated.

Sally Haslanger's account is the most well-known and widely discussed theory of this sort. As I mentioned earlier, Haslanger does not attempt to analyze our ordinary concepts of *woman* and *man*, but rather to provide ameliorative definitions of these terms. To be a woman on her view is (roughly) to occupy a subordinate social position by virtue of being regularly perceived (rightly or wrongly) to have a female body. Or, more fully:

> S is a woman if, and only if, (i) S is regularly and for the most part observed or imagined to have certain bodily features presumed to be evidence of a female's biological role in reproduction; (ii) that S has these features marks S within the dominant ideology of S's society as someone who ought to occupy certain kinds of social position that are in fact subordinate (and so motivates and justifies S's occupying such a position); and (iii) the fact that S satisfies (i) and (ii) plays a role in S's systematic subordination, i.e., along some dimension, S's social position is oppressive, and S's satisfying (i) and (ii) plays a role in that dimension of subordination.

By contrast, to be a man is to occupy a dominant social position by virtue of being regularly perceived to have a male body. (The fuller definition corresponds to her definition of "S is a woman", with changes in just the places you'd expect them.) Haslanger does not include definitions for alternative gender categorizations—e.g., *gender fluid*, *agender*, *nonbinary*, and so on. Her account is not in principle committed to a binary conception of gender; but whether her account can accommodate other gender categories depends on whether there are social positions corresponding to those categories.

One consequence of this account, which Haslanger herself points out, is that the proper goal for those promoting social equality is not,

strictly speaking, to bring about social equality *for men and women as such*, but rather to make it the case that nobody counts as a man or a woman. At first glance, this might seem an odd, and even problematic, goal, one that might amount to working toward the goal of making males non-masculine and females non-feminine. But to raise this as an objection is simply to presuppose that gender is something other than what Haslanger thinks it is. If one takes seriously the idea that gender categories are positions in a system of domination and subordination, the goal makes good sense. On this understanding of gender, the practices whereby people are assigned to gender categories are inherently oppressive, and so those concerned for equality will naturally seek to reform society in ways that do away with the entire system of gender.

Note, too, that if Haslanger's account is correct, gender categories are constructed by a particular social ideology together with the observations and imaginings of the people within the society that embraces the ideology. This raises a host of additional questions in social metaphysics. What does it mean for a *society* (as contrasted with an *individual*) to embrace an ideology? Is a society a group mind or corporate person, such that it can have thoughts and beliefs in the way that individual minds or persons do? Or might we instead endorse a deflationary account, according to which a society believes something just so long as some salient percentage of its members believes it? And what exactly constitutes a society. Are the US and UK two separate societies, or just parts of one Western, or English-speaking society? If, as seems plausible, they are separate societies, how fine-grained do societies get? Is your home state a society? Your county? Your city? Your family? Are churches, mosques, synagogues, and such societies? Is an entire religious tradition a society? And if we answer yes to some of these questions, does that mean (for example) that one might be a woman relative to US society but not relative to one's family? Obviously we cannot pursue these questions here; but I raise them just to remind the reader that there is a *lot* more worth exploring in social metaphysics than I am able to treat in a single chapter.

In the section on mind dependence, I indicated that views according to which gender is socially constructed allow for relativization, or contextual variation in a person's gender membership—that is, one

might belong genuinely to one gender in some contexts and another gender (or no gender) in other contexts. This context sensitivity arises in Haslanger's account in two distinct ways. First, there can obviously be variation among societies and cultures in what social position (if any) a person occupies by virtue of how their body is perceived in that culture; and there may well also be variation in how certain bodies are perceived. This will straightforwardly allow for someone to be a woman (or man) in one cultural context but not in another. But, second, insofar as Haslanger's account provides an *ameliorative* definition, we already know that it does not capture the socially dominant meanings of the words "man" and "woman". (The very project of providing an ameliorative definition is predicated on the idea that the dominant meaning is not serviceable for one's purposes.) As a result, one might understand Haslanger's view as allowing that a socially dominant female like the Queen of England *is* a woman relative to contexts where the dominant meaning of "woman" is in play, but *is not* a woman (by virtue of her privileged social position) relative to contexts where "woman" means what Haslanger proposes. Haslanger herself seems not to want her view interpreted this way. She allows, for example, that a woman might not *function as such* in certain contexts; and this seems to suggest that she thinks that whether one *is* a woman does not vary from context to context within one's own society or culture. But it is, nonetheless, a possible view one might hold.

An alternative to characterizing gender as a position in a social hierarchy, and the third "main option" I will discuss, is to characterize it simply as a social *property*—one that may or may not locate a person in a social hierarchy. This is the view defended by Ásta in *Categories We Live By*. According to Ásta,

> a social property of an individual is one that one has because of something about other people ... it is a property that someone has conferred on them by others. This property is a social status consisting in constraints on and enablements to the individual's behavior in a context.

In saying that social properties *are conferred on us by others*, Ásta does not mean to suggest that acquisition of a social property is always a matter of someone else's decision. Rather, a social property is conferred under certain conditions by attitudes or actions that consciously or

unconsciously aim to track some underlying "base property". In the case of a strike in baseball, for example, the property *being a strike* is conferred by the declared beliefs of the umpire in the context of a baseball game; and the umpire's beliefs aim to track the ball's trajectory, the base property. In the case of gender, Ásta thinks that gender properties are conferred in different contexts by the beliefs and attitudes of people who have what you might call the relevant kind of "social authority", or (as Ásta puts it) "standing", in that context—people who enjoy social dominance within the context, for example, or people who have a certain kind of institutional authority (e.g., doctors, judges, etc.). These beliefs and attitudes aim to track certain base properties of the individuals upon whom gender is conferred. But, according to Ásta, *which* base properties are tracked varies from context to context. In some contexts, gender is conferred on the basis of one's reproductive role; in other contexts it is conferred on the basis of one's social role, or perceptions about one's body, etc. In this way, the context sensitivity of gender assignment—again, not just gender *judgments*, but possession of gender *properties*—is built right in to Ásta's account.

Ásta's *conferralist* account of social properties is not the only available theory of how social properties are acquired. Ásta contrasts conferralism with two other accounts of social properties: the *constitution* account, and the *response dependence* account. Roughly, the constitution account maintains that social properties are constituted by what Ásta calls the base property. Thus, a strike is simply a ball flying along a certain trajectory in the context of a baseball game, regardless of whether the umpire sees it that way; and having a particular gender is just having the corresponding base property—a particular sex, or a particular social role, or whatever. The response dependence account maintains that having a social property is just a matter of inducing the right kind of response in the right kind of people. So a strike will be any thrown ball that is treated or responded to as a strike by the umpire, regardless of the ball's actual trajectory, what the umpire believes, or what, if anything, the umpire is aiming to track. Likewise, having a particular gender is a matter of being someone who is generally treated or responded to as belonging to that gender, regardless of one's actual sex or social role, other people's beliefs about one's sex or social role, and so on. I won't explore these other accounts in further detail; but it is important to note that they

correspond to two other types of social-property account of gender that one might try to defend.

Ásta's account has some important advantages. It provides an intuitively plausible account of how we get our gender properties. It also allows that, and explains how (to use one of Ásta's examples), one and the same person can experience themselves as one of the guys at work, one of the women at a gathering for their grandma's 80th birthday, and neither a man nor a woman at a local bar. Returning to the baseball analogy, it is as if each of these different contexts is like a new baseball game with both a different umpire and a different strike zone. Just as one and the same trajectory might count as a strike in one game but not the other, due to differences in the umpires and the base properties they are aiming to track, so likewise one and the same individual might count as a woman in one context but not another due to differences among the relevant "gender umpires" and the base properties they are (consciously or unconsciously) aiming to track.

But Ásta's account also has consequences that many will find objectionable. For example, it implies that self-identification as a woman or man in a context is neither necessary nor sufficient for being a woman or a man in that context. One is not always one's own "gender umpire". Rather, it may sometimes be other people or social institutions, rather than oneself, who have the standing, or authority, to assign gender. Moreover, and for the same reason, Ásta's account allows for the possibility that one can be *mistaken* about whether one is a man, a woman, or a member of some other gender category; and therefore one will not necessarily be *misgendered* by people who disagree with one's own gender self-assessment. These consequences are not unique to Ásta's conferralist account; they seem, for example, to be consequences of Haslanger's view as well. But for some philosophers these consequences will be decisive reasons to reject Ásta's theory of gender.

This brings us to the third and final alternative—the idea that gender is a *self-conferred identity*. (I'll call this "the SCI account".) As we have seen in an earlier chapter, the branch of metaphysics that deals with puzzles about *identity over time* is most centrally concerned with the concept of *numerical identity*. But the notion of identity in play here is different from that (and sometimes enters confusingly into

the literature on personal identity). Roughly, an identity in the sense I have in mind here is something central about *who you are*. I am a tennis player; but being a tennis player is not really an "identity" of mine in this sense because it is not all that central to who I am. By contrast, being a philosopher, being white, and being a Christian *are* identities of mine. These category-memberships are quite central to who I am. Some, like my professional and religious identities, are ones that I chose for myself; others, like my racial identity, are ones that I was born into. Still others, like being disabled (which is not an identity of mine), might be neither hereditary nor chosen; they might instead be conferred upon a person by way of the interaction of (i) differences between one's own body and the bodies of most other people (e.g., differences in mobility, perceptual capacities, etc.), and (ii) the wide variety of social factors that explain why people sharing one's own particular differences have reduced access to public spaces, certain kinds of jobs, or other social goods. For those who think that gender is a self-conferred identity, one belongs to a gender category if, and only if, one sincerely self-identifies with that category.

Given that gender is *self*-conferred on the SCI account, one might wonder whether it can properly be called a *social* constructionist account. Admittedly, the label "social" is an uneasy fit in light of the fact that the theory implies that no *individual's* gender is constructed by social ideologies, institutions, or conventions. But remember that we have defined social constructionism not as the view that some kinds are constructed by *society*, but rather as the view that *some social kinds are constructed*; and gender is still a constructed kind on this view. Moreover, the very idea of gender, as well as the social norms, stereotypes, and so on in accord with (or in resistance to which) each person conceptualizes, performs, and thereby constructs their gender seem clearly to be not only constructed, but constructed *by society*. So there is ample reason to classify the SCI account as a social constructionist view.

One might also wonder whether the SCI account might be appropriately treated as a variant of Ásta's conferralist view. Granted, the SCI account shares with Ásta's account the basic idea that gender is conferred rather than being constituted by a particular base property or by the responses that the gendered individual induces in others.

But there the similarities end; and so it seems that the SCI account is best treated as a distinct view in its own right.

The SCI account is the predominant view in the field of Women's, Gender, and Sexuality Studies. Furthermore, of the characterizations offered so far, it is the one that is most friendly to trans and queer, or non-binary identities. All of the previous theories allow for one's gender to be fixed by facts about one's birth-sex, independently of whether one actually identifies with that gender; but to have a trans identity or queer, or non-binary gender identity is, in part, to deny that one's gender is fixed by the sex that one is assigned at birth. Thus, all of the other theories so far considered allow for a tension between the "facts" about a person's gender according to that theory, and that person's first-person judgments about their gender. As Talia Mae Bettcher points out, however, the view assigns first-person authority to gender judgments. That is, since everyone has privileged access to the facts about what gender category they themselves sincerely identify with, each individual person is authoritative with respect to judgments about their own gender. For many, this will be a decisive advantage of the SCI account over the others we have considered thus far.

I think the most obvious challenge faced by the SCI account as I have so far presented it, however, is answering the question of what, exactly, gender *is*. Yes, it is an identity; and yes, we know the various factors (biological, social, psychological, performative, and so on) that enter into a person's gender identity. But what *kind of thing* is an identity? Is it, for example, a belief, an attribute, an attitude, a personality type, a social role, or something else? Saying that it is "something central to who you are" as I did above doesn't really answer that question, and I am not aware of any proponent of the SCI account who has given a fully detailed answer. This is not, of course, to say that there are no explanations of the relevant sense of *identity* whatsoever in the literature; but the SCI account requires a particular kind of explanation—one that will be amenable to the idea that people confer their gender identity upon themselves, and in a way that gives them first-personal authority over their gender assignment.

Borrowing some words from psychologist Kate McLean, I suggest thinking of an identity as a mental construct—perhaps a representation or concept of oneself, or an internalized narrative about

oneself—that brings "a sense of coherence and integration to one's life, allowing [one] to perceive a sense of continuity through time". On this way of thinking, gender is neither an attribute nor a social position; it is, rather, a kind of self-interpretation. It is perhaps tempting to object that thinking of gender in this way turns it into something that makes little difference in our lives—it is just an aspect of how we think about ourselves. But there is a vast and growing literature in both philosophy and psychology according to which the narratives we tell about ourselves, and the identities that correspond to or are constituted by them, make a substantial difference in our thoughts and actions, and in what we can and cannot do. If this is right, then there may be ways of developing the SCI account so that it captures some of what is attractive about other accounts—e.g., Asta's insight that gender properties are attended by certain kinds of *constraints and enablements*, as well as the insight shared by social position accounts that gender properties tend to imply certain kinds of social position. But at this point we have only a promissory note; for, of the accounts of gender currently available, the SCI account, for all its popularity, seems to be the one that has received the least philosophical development.

FURTHER READING

For those interested in constructivism, I recommend Nelson Goodman, "Words, Works, and Worlds", reprinted from his *Ways of Worldmaking* (Indianapolis, IN: Hackett, 1978) in *Arguing about Metaphysics*, Hilary Putnam, *Ethics without Ontology* (Cambridge, MA: Harvard University Press, 2005), and Alvin Plantinga, *Twin Pillars of Christian Scholarship* (Grand Rapids, MI: Calvin College and Seminary, 1990). Michael Devitt's *Realism and Truth*, 2nd ed. (Princeton: Princeton University Press) is also an excellent resource for discussion of realism and anti-realism generally.

On gender, I mentioned Simone de Beauvoir's claim that one is not born, but rather becomes a woman. This is from her landmark work, *The Second Sex*, tr. Constance Borde and Sheila Malovany-Chevallier (New York: Alfred A. Knopf, 2010). I also quoted from the following works: Sally Haslanger, "Gender and Race: (What) Are They? (What) Do We Want them to Be?", *Noûs* 34 (2000): 31–55; Riki Wilchins, *Queer Theory, Gender Theory: An Instant Primer* (New York: Magnus Books, 2014); and Kate McLean, *The Co-Authored Self: Family*

Stories and the Construction of Personal Identity (Oxford: Oxford University Press, 2015). The idea of gender *performativity* which made an appearance toward the end of the chapter is from Judith Butler's *Gender Trouble* (New York: Routledge, 1990). The views on gender of both Judith Butler and Luce Irigaray (see, e.g., *Speculum of the Other Woman* [Ithaca, NY: Cornell University Press, 1985] and *This Sex Which Is Not One* [Ithaca, NY: Cornell University Press, 1985]) have been tremendously influential; but I have omitted discussion of them here because their work lies well outside the domain of analytic metaphysics, with which the present book is primarily concerned.

On the metaphysics of gender in general, I strongly recommend—and my discussion here has been significantly informed by—Ásta, *Categories We Live By* (Oxford: Oxford University Press, 2018); Elizabeth Barnes, "Gender and Gender Terms", *Noûs* 54 (2019): 704–730; Talia Mae Bettcher, "Trans Identities and First-Person Authority", in Laurie Shrage (ed.), *You've Changed: Sex Reassignment and Personal Identity* (Oxford: Oxford University Press, 2009), 98–119, and "Trans Women and the Meaning of Woman", in Nicholas Power et al. (eds), *The Philosophy of Sex*, 6th ed. (Lanham, MD: Rowman & Littlefield, 2013), 233–250; and Robin Dembroff, "Escaping the Natural Attitude about Gender", *Philosophical Studies* 178 (2021). Dembroff's paper is a reply to Alex Byrne's "Are Women Adult Human Females?", *Philosophical Studies* 177 (2020): 3783–3803. I have learned a lot from Charlotte Witt's fascinating *The Metaphysics of Gender* (Oxford: Oxford University Press, 2015), which defends a form of essentialism that does not fall prey to the usual concerns about that sort of view. I recommend it to readers, but I did not discuss it here because of the complexity of the view.

In discussing the harms of gender essentialism, I mentioned Jean Jacques Rousseau, Elizabeth Spelman, and Mary Wollstonecraft. The works I have in mind, as well as a variety of others (both historical and recent) can be found in Mary B. Mahowald (ed.), *The Philosophy of Woman*, 3rd ed. (Indianapolis, IN: Hackett, 1994).

Finally, I said that I would recommend readings on three other important topics in social metaphysics: race, disability, and sexual orientation. On race, a good place to start is Joshua Glasgow et al., *What is Race? Four Philosophical Views* (Oxford: Oxford University Press, 2019). On disability, I strongly recommend Elizabeth Barnes, *The Minority Body* (Oxford: Oxford University Press, 2016). Finally, on the metaphysics of sexual orientation, interested readers should see Robin Dembroff's groundbreaking paper, "What is Sexual Orientation?", *Philosophers' Imprint* 16, no. 3 (2016): 1–27.

12

METAPHYSICS AND ITS CRITICS

I mentioned in the first chapter of this book that many people start asking metaphysical questions in early childhood. It is entirely natural to wonder even at a very young age about the nature of time, of the soul, of our own free agency, and so on. What might seem a bit surprising upon reflection, however, is that the drive to answer these questions persists well into adulthood. Why in the world would anyone (but a child) think that human beings are capable of answering questions that pertain to matters not only outside the grasp of our own sensory apparatus but beyond the reach of science altogether? Why think that human beings have any hope at all of answering *non-empirical* questions about what exists or about the fundamental natures of things?

The fact that metaphysics concerns itself with non-empirical questions about the character and makeup of the world has made many philosophers uneasy with the whole enterprise. The sciences enjoy a great deal of respect as fields of inquiry, and many think that the methods of science and those methods alone are the tools by which we ought to build our theories about the world. ***A priori*** theorizing about the world—the sort of theorizing that requires no lab-equipment or experimental apparatus, but just a rocking chair, a working brain, and a good chunk of time free for thinking—has long been viewed with

skepticism. According to many philosophers, metaphysical theorizing is just idle tale-spinning. In this closing chapter, I want to consider and address a couple of different variations on this sort of objection.

AGAINST METAPHYSICS

One of the most well-known expressions of this sort of negative attitude toward metaphysics comes from David Hume (1711–1776). In section XII of his *Enquiry Concerning Human Understanding*, Hume writes:

> If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, Does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.

Hume's attitude is characteristic of those in the **empiricist** tradition—those who treat empirical theorizing of the sort found in the sciences as the only way of acquiring reliable information about the world. Likewise, contemporary empiricists have had little better to say about metaphysics. Rudolf Carnap (1891–1970), for example, complained in his article "The Elimination of Metaphysics Through Logical Analysis of Language" that metaphysicians are like "musicians without musical ability", the point being that metaphysical theories are devoid of cognitive content and (unlike good music) not even endowed with aesthetic value.

   What we have here are two different, but closely related concerns. The Humean concern can be fleshed out as follows. We have good reason to believe that the methods of science are reliable sources of information about the world. Those methods have enabled us to make striking and tangible improvements in our ability to predict and control our environment, and also in our ability to develop useful and sometimes downright amazing pieces of technology. Matters are different in metaphysics, where the staple mode of inquiry is nothing more than *a priori*, intuition-based theorizing. Notably, metaphysics as a discipline doesn't make discernible progress in the way that the sciences do. In metaphysics, we find widespread and persistent theoretical disagreement, very few problems that the community of

metaphysicians take to have been decisively solved, and no clearly measurable track record of success whatsoever. *One* conclusion a person might draw from the difference is this: Unlike the methods of science, the methods of metaphysics show no sign of being reliable ways of increasing our knowledge about the world; thus, there is some reason to suspect that metaphysical theories are nothing more than "sophistry and illusion".

Carnap also privileges empirical investigation over *a priori* theorizing. But from his point of view, the problem isn't simply that the methods of metaphysics are unreliable. On his view, the "questions" discussed by metaphysicians don't even rise to the level of being meaningful. Underlying this view is the idea that a question is meaningful only if it can be answered somehow by appeal to observation, empirical theory, or the meanings of the words in the question. In other words, the meaningful questions are the ones that science can answer and the ones that can be answered by appeal to **analytic** truths—sentences like "bachelors are unmarried", which are commonly said to be purely conceptual truths, expressible by sentences that are "true by definition", or such that their truth is grounded solely in the meanings of the terms involved. If the questions can't be answered in one of these ways, then they are meaningless; and if they can be so answered, then the *correct* answers are just the answers given by observation, empirical theory, or the relevant definitions. So, for Carnap, to ask as the metaphysician does whether the entities referred to in our empirically supported theories *really* exist is silly. Either the question is meaningful and already settled by whatever empirical evidence we have for the theory, or else it is meaningless and so a pseudo-question.

Let us consider an example. In mathematics we have axioms, definitions, and theorems that imply apparent existence claims like "there are two prime numbers between the numbers four and eight". Definitions explicitly provide the meanings of some of our mathematical terms; axioms implicitly supply meanings for other terms; theorems express the logical consequences of the axioms and definitions. Now suppose someone asks, "are there *really* two *things*—prime numbers—between two other things called 'the number four' and 'the number eight'?" As Carnap sees it, this question can be taken in one of two ways. It can be taken as a question *internal* to the

mathematical enterprise, which is to say that it can be taken as nothing more than an ordinary mathematical question. Taken that way, the question is perhaps a bit puzzling, but the answer is obvious: yes, of course there are two prime numbers between four and eight! The answer follows straightforwardly from the axioms, definitions, and theorems of our mathematical theory. It is, in other words, an analytic truth. Thus, the question is meaningful, and so likewise with other internal mathematical questions.

Alternatively, our question can be taken as an *external* question, asking from a point of view outside the mathematical enterprise whether the existential statements implied by mathematical claims—again, claims like "there are two prime numbers between four and eight"—*really* commit us to believing in things like numbers. That question, Carnap says, is meaningless. We can give it meaning by taking it as an indirect way of asking whether it is a good idea to continue with our current systems of mathematical axioms, definitions, and theorems. This latter question is meaningful because it can be settled by re-examining the place of mathematical discourse in our overall body of scientific theory. We might ask, for example, whether our scientific theories might be more useful to us, or more successful at making verifiable predictions, if we were somehow to stop using current mathematical theories. These are empirical questions, and so they are meaningful. But to take the question of whether there really are numbers any other way, he thinks, makes no sense. For, in general, a question is intelligible only under the supposition that its answer is an analytic truth or a truth discoverable by way of empirical investigation; and this is precisely the supposition we rule out when we take ourselves to be asking an "external question".

Carnap's concern is motivated by strong assumptions about the conditions under which a question is meaningful. It also seems to assume that we can sensibly distinguish between analytic and non-analytic truths. These assumptions are controversial. But "neo-Carnapian" critics of metaphysics have raised a very similar objection against metaphysics that does not depend on Carnap's assumptions. The objection starts with the observation that, regardless of *how* we know that there are two prime numbers between four and eight, the fact is that we *do* know it. It is an indisputable mathematical fact that

there are two prime numbers between four and eight; and the following argument form is obviously valid:

(1)   There are two Fs between x and y.
(2)   Therefore: There are (at least) two Fs.
(3)   Therefore: Fs exist.

Thus, we can simply deduce from obvious, indisputable facts the conclusion that there are numbers. And the same might be said about questions like, "Are there really chairs?", or any of the other existence questions that metaphysicians like to ask.

Why, then, do philosophers continue to argue over questions like "are there prime numbers?", given that they have such easy answers? A common diagnosis is that parties to metaphysical disputes are simply talking past one another. In other words, each party is using relevant words differently from the way the other party uses them, and they are also using words differently from the way ordinary speakers use them. So, for example, when philosopher A says "there are no prime numbers" and philosopher B disputes their claim, A is using at least one of the terms in the disputed claim—probably either "there are" or "numbers"—differently from how B is using them, and both A and B are using at least one of the terms differently from how they are used by ordinary speakers, for whom the truth of the claim is obvious. Notice that even where there are metaphysical disputes about the existence of things like tables or numbers, there is no real dispute about the "underlying facts" that lead us to say things like "there are two prime numbers between four and eight" or "there are chairs". Most philosophers who would deny that there are prime numbers *wouldn't* deny claims like "five is prime", "seven is prime", or "five and seven are between four and eight". Likewise, philosophers who deny that there are chairs will not hesitate to say there is matter arranged chair-wise. So many philosophers think that all of this evidence together suggests that the questions of metaphysics, even if not strictly meaningless, are at least not substantive.

The upshot of this third objection, then, is that metaphysicians in dispute with one another are like two people arguing over the question whether the crust leftover from a piece of pizza itself counts as

pizza. One says that something is pizza only if it has sauce, crust, and cheese; the other says that, since the crust was part of the original pizza, one is still eating *pizza* even when one is down to eating nothing but the crust. There is obviously no real or interesting dispute here; it is just a silly difference about how to use the word "pizza". Settling the debate isn't a matter of looking into the fundamental nature of pizzahood, or of anything else. It is, rather, just a matter of taking a closer look at our language, or stipulating a particular way in which the term "pizza" is to be used henceforth. So likewise, say the objectors, with metaphysical disputes about the existence of tables, numbers, and the like.

IN DEFENSE OF METAPHYSICS

We now have three related objections against metaphysics on the table. First: the methods of metaphysics are not reliable means of discovering truths about the world; second, the questions of metaphysics are meaningless; and, third, even if the questions of metaphysics aren't strictly meaningless, they are at least non-substantive, giving rise to mere verbal disputes. In light of these sorts of objections, it is easy to see why many people think that academic discussion of metaphysical topics is neither serious nor important.

Before considering answers to these objections, we should first observe one difference between the first objection and the second two. If the first objection is sound, then the *whole field* of metaphysics is indicted. Matters are less clear with the second two. We could perhaps acknowledge that certain metaphysical questions are meaningless, or that certain disputes in ontology are mere verbal disputes, without necessarily calling the whole field of metaphysics into question. For the fact is metaphysicians examine and critically evaluate some of the most existentially important beliefs that human beings ever hold—beliefs that lie at the very heart of our conception of ourselves and our commonsense ways of thinking about the world. It matters very much to us whether we are free in a way that would allow us to be genuinely responsible for our actions, whether we might possibly survive death, whether there might be things— perhaps even intelligent and powerful things—beyond what we find in the material world, etc. It would take substantial argument to show

that all of these questions are meaningless, or that our disputes on all such topics are mere verbal disputes.

But now what should we think of our three objections? Let us begin with the Humean objection. There are at least two ways of responding. First: the starting point for the objection is the fact that metaphysicians rely on intuitions (again, intellectual experiences of the sheer obviousness or necessary truth of various claims) in building their theories. But it is notoriously hard to produce an *argument* against the reliability of intuition as a source of evidence without *relying* on intuitions. Is persistent disagreement and ongoing failure to reach decisive solutions to the discipline's central puzzles and problems really evidence of general lack of progress? Is lack of *the sort of progress that one finds in the sciences* really evidence that the methods of a discipline are unreliable? These are philosophical questions, answerable only by way of the very methods on which metaphysicians rely—methods that include appeals to intuition as a source of evidence. Our answers to these questions will depend significantly on our intuitions about the nature of *progress* and about the conditions under which we ought to believe that a method of inquiry is unreliable. So if one answers both questions affirmatively (which one must do in order to defend the main premises of the Humean objection), one successfully indicts intuition; but one also thereby indicts the very methods by which one reached the affirmative answers. Clearly that won't do. If, on the other hand, one answers the questions negatively, then the Humean objection cannot get off the ground.

In light of the reply just given, many contemporary metaphysicians have felt content simply to ignore objections arising out of the empiricist camp, at least until those objections can be made precise enough to be persuasive. It might be tempting to think that this reaction comes to nothing more than sticking one's head in the sand. Even if we grant that there is something self-defeating about a *philosophical* objection to the methods used in one of the core sub-fields of philosophy, can't we at least acknowledge that it's not *obvious* that the methods of metaphysics are reliable? And if so, shouldn't metaphysicians be working hard to show that their methods are reliable? And isn't their failure to do so itself rather telling?

In a word, no. Indeed, the "head in the sand" response is exactly the reaction that many empiricists themselves have toward more general

skeptical complaints—complaints that challenge the evidential value of sense perception, for example. When skeptics demand evidence for the reliability of our senses, empiricists generally take themselves to be justified in ignoring the demand. You can call it sticking your head in the sand; but to do so presupposes what most philosophers would not grant—namely that the skeptic has issued a legitimate challenge that demands a response.

Contrary to this presupposition, most philosophers now think that we are entitled simply to trust our senses, and that the demand for evidence that demonstrates them to be reliable cannot be met. If that is right, then shouldn't we be entitled simply to trust our rational intuitions as well? Many have thought so.

In this vein, consider what Thomas Reid (1710–1796) says in his *An Inquiry into the Human Mind*:

> The sceptic asks me, Why do you believe the existence of the external object which you perceive? This belief, sir, is none of my manufacture; it came from the mint of Nature; it bears her image and superscription; and, if it is not right, the fault is not mine: I even took it upon trust, and without suspicion. Reason, says the sceptic, is the only judge of truth, and you ought to throw off every opinion and every belief that is not grounded on reason. Why, sir, should I believe the faculty of reason more than that of perception? – they came both out of the same shop, and were made by the same artist; and if he puts one piece of false ware into my hands, what should hinder him from putting another?

The point here is that reason (which includes rational intuition) and sense perception ought to be treated as on a par—at least initially, unless and until we get evidence that one or the other is untrustworthy. Why? Because whatever processes gave rise to the one almost certainly also gave rise to the other. Reid, of course, is defending the trustworthiness of sense perception, but exactly the same sort of speech might equally be made in defense of intuition. To be sure, those offering Reidian speeches on behalf of intuition-based theorizing must grant that there is more disagreement and less "progress" in metaphysics than in science. But many agree that even once this difference is accounted for, the alleged reasons for abandoning metaphysics are far from compelling.

A second reaction to the Humean objection has been to take on the empiricist preference for empirically based theorizing, along with empiricist skepticism of appeals to intuition, and to try to do metaphysics in a way that is roughly continuous with science, going beyond science as little as possible. Those who identify with the tradition of *philosophical naturalism* (roughly, the idea that only the methods of science can be trusted by default, without some evidence proving their reliability) manifest a deep skepticism about appeals to intuition. So they confine themselves to a sort of metaphysics that aims simply to fill explanatory gaps in our scientific theories and to draw out some of the interesting logical consequences of those theories. This reaction concedes a lot of territory to the critics. But, if sound, it still leaves an important place in the overall project of human inquiry for the practice of metaphysics.

Initially, this way of thinking about metaphysics might seem puzzling. It looks like the philosophical naturalist agrees with Carnap in thinking that metaphysicians shouldn't be looking to answer questions that can't be answered by science. The naturalist wants to go beyond science as little as possible; and, though it is surely a fine thing for metaphysicians to spend their time filling gaps in scientific theories and drawing out interesting logical consequences, it is not as if working scientists couldn't do those things for themselves. So, what *distinctive work* is left for the metaphysician to do? And why isn't the naturalist just a Carnapian in disguise?

Remember that Carnap's objection to metaphysics rested in part on a sharp distinction between analytic and non-analytic truths, and also on a particular view about the conditions under which a question or statement might be meaningful. But the distinction between analytic and non-analytic truths is now widely regarded as suspect; and once that distinction is blown, Carnap's views about the conditions under which a question or statement is meaningful are also in trouble. After all, claims of mathematics and logic (to take just two examples) are meaningful, but not directly empirically verifiable. For Carnap, this wasn't a problem because these claims could plausibly be regarded either as analytic truths or as truths derivable from some combination of analytic and empirically verifiable truths. Without the analytic/non-analytic distinction, *this* reason for thinking that the claims are meaningful disappears. But it is hard to see any other way

of guaranteeing their meaningfulness that wouldn't also open the door to saying that lots of other empirically unverifiable claims—including claims of metaphysics—are meaningful as well. These criticisms of Carnap were pressed most famously and strongly by W. V. Quine (1908–2000). Quine shared Carnap's commitment to science and to certain other principles of traditional empiricism. But he nevertheless thought that questions in ontology were meaningful, and that philosophers could give them meaningful answers. This is one crucial difference between the naturalist and the Carnapian.

The difference between the Quinean and Carnapian perspectives on metaphysics depends partly on differences in their conception of how empirical evidence supports the various things that we believe. Like a lot of us, Carnap took it for granted that there is a reasonably sharp boundary, with no substantial middle ground, between claims that are empirically testable and claims that aren't. As he saw it, the empirically testable ones were the meaningful and scientifically respectable claims; the rest were either analytic truths or meaningless. Quine, however, thought that empirical evidence can provide support for a claim by supporting an *overall body of theory* that includes the claim in question as an integral part. So, for example, principles of mathematics and logic can be seen as *empirically* supported even though they obviously aren't directly testable. This is because they are important parts of an overall web of belief which includes our best scientific theories, and which enjoys, as a whole, a large measure of support from direct observation, experimental tests, and other empirical methods. For Quine, the idea that we can go through our total set of beliefs and sort them neatly into boxes marked EMPIRICALLY VERIFIABLE, ANALYTIC, and MEANINGLESS is wholly misguided. What we have instead is a web of beliefs which is continually revised in light of our experiences, and which is more or less confirmed *as a whole* by the way in which the beliefs together cohere with one another and with our total experience.

Quine, unlike Carnap, is prepared to acknowledge that there are many scientifically respectable claims that are not empirically testable. Such claims, he says, fill in some of the gaps left by our scientific theories and sometimes lead to further hypotheses that can be tested. At any rate, the acceptable ones do this; and many of the claims made by metaphysicians can be seen as fitting this description. Moreover, since

there is no sharp line between claims that are testable and claims that merely fill gaps in our theories, the boundary between science and metaphysics is likewise unclear. But that does not matter; for Quine has no interest in drawing sharp lines between philosophy and science. Quite the contrary: as a naturalist, he sees the disciplines as continuous with one another.

Thus far I have been explaining *why* Quine thinks that philosophers can answer questions in ontology. But that still leaves us with the question of *how* such questions are to be answered—especially since Quine himself is no fan of intuition-based theorizing. The answer is that we look to our best scientific theories and ask ourselves, "Which of the entities apparently referred to by these theories *have* to exist in order for the theory to be true?" So, for example, suppose our theory says the following:

(T)   The average family has a mother, a father, and 2.2 children.

The ontologist might ask: "Do there *have* to be such things as average families, and mothers, and fathers, and children in order for this theory to be true?" A fairly natural answer is to say that there *do* have to be mothers, fathers, and children, but there *doesn't* have to be any such thing as the average family. Why? Because it is fairly easy to *paraphrase* the sentence in a way that gets rid of apparent reference to average families, whereas it would be much harder to paraphrase away all apparent references to mothers, fathers, and children.

We can put all of this a bit more precisely as follows. Philosophers generally assume that claims like "chairs exist" or "there are chairs" are logically equivalent to "something is a chair". Now, most commonsense and scientific theories (no matter how we express them) will include or logically imply a variety of existential claims—again, sentences that have this form:

(E)   Something is an F,

where "F" stands in for some kind-term like "chair" or "number". Quine thinks that if a theory implies a sentence with the same form as E, then that theory is *committed* to whatever must exist in order for that sentence to be true. So, for example, if T really implies that

something is an average family, then our theory T is committed to the existence of average families. The question for the ontologist, then, is whether T really does imply that something is an average family. Most of us would say that it does not. According to Quine, the way to be sure that it does not is to try to express T in a different way—to paraphrase it. If we find that T can be stated without using terms that refer to average families, then we are entitled to say that T does not imply that something is an average family. If we fail in our efforts to find a suitable paraphrase, then we must admit that T is committed to average families after all.

It is easy to see how this way of thinking about ontology will enable us to ask and explore a wide variety of interesting questions. Our best mathematical theories tell us that there are two prime numbers between four and eight. So one apparent implication of those theories is E1:

(E1)   Something is a number.

Our best biological theories seem to tell us that spiders and insects share several features in common. So one apparent implication of those theories is E2:

(E2)   Something is a feature.

Are E1 and E2 *genuine* implications? For Quine, that all depends on whether there are acceptable paraphrases of the prime number sentence, or of the spider–insect sentence, that don't make reference to numbers or features. Are there such paraphrases? That is a hard question. Moreover, it is not at all clear that we would have to rely on anything as suspect as rational intuition in order to answer it. In any case, according to Quine, answering that question is precisely what it takes to find out whether our best confirmed theories are committed to things like numbers and features; and isn't it obvious that we should believe in the things to which our best theories are committed?

Before moving away from the Humean objection which has been our focus thus far, I want to close this discussion of Quine by noting

that, although Quine is a naturalist, Quine's approach to ontology has been embraced wholeheartedly by non-naturalists as well. The basic idea underlying Quine's approach to ontology is entirely commonsensical: If you have a theory that seems to imply that something is an F, you seem to face a choice between three options when pressed on the question as to whether you really believe in F's: (a) retract the claim that seems to imply that there are F's; (b) say that the implication is only apparent, and supply a paraphrase that doesn't refer to F's; or (c) accept the consequence of your theory and admit that you believe in F's.

We have now finished dealing with the Humean objection. Moreover, we have already seen along the way how the second, Carnapian objection might be dealt with. Thus, we shall now move on to the third objection: the neo-Carnapian concern that metaphysical debates are, at best, nothing more than mere verbal disputes.

As with the Humean objection, there are at least two ways of responding to the neo-Carnapian objection. One way is to try to stake out a middle ground between Carnap and Quine. One might take on board the idea that metaphysical questions are, in large part, questions about the meanings of our words but nevertheless maintain that at least some such questions are subject to genuine and substantive philosophical dispute. This middle position is defended by Amie Thomasson. Thomasson's views are generally characterized as neo-Carnapian because of obvious affinities that her views bear with those of other neo-Carnapians. Despite this, I still think that it is fair to characterize her as offering a response to more extreme neo-Carnapians who think that all or virtually all metaphysical disputes are mere verbal disputes.

In the concluding chapter of her book, *Ordinary Objects*, Thomasson observes that contemporary philosophers tend to think of metaphysics as analogous to science. Theories are evaluated on the basis of criteria like simplicity and explanatory power; and there is every expectation that inquiry will ultimately lead to theories that force us to *revise* many of our previously held "pre-theoretical" beliefs. After all, there is no reason to think that our overall body of commonsense beliefs will be respected by the simplest and most explanatory theory. But, she argues, one of the lessons of her book is that this way of thinking about metaphysics is incorrect: Distinctively metaphysical

questions are to be answered through a kind of conceptual analysis—an activity pursued via reflection on the meanings of our words—rather than through methods of the sort we find in science.

So, for example, suppose we are raising metaphysical questions about whether there are numbers, or tables. On Thomasson's view, the way to address these questions is to analyze our concepts of *number* and *table* and to determine the conditions under which those concepts would apply to something. Doing this is difficult, and the results may well be controversial. So, on her view, there is no reason to think that debates about the application conditions for our concepts of *number* and *table* will be meaningless, trivial, or merely verbal. But, she says, once a view has been reached about the application conditions for these concepts, the answers to questions like "Are there numbers?" or "Are there tables?" are straightforward. We simply look to see whether the application conditions are fulfilled.

Nevertheless, Thomasson does think that many of the "typical" debates in metaphysics *are* mere verbal disputes. The reason is that philosophers are not typically asking after the application conditions of particular terms like "number" or "chair"; they are, rather, asking more general questions about what kinds of *things* or *objects* there are and, on Thomasson's view, the terms "thing" and "object" are subject to a variety of different uses. The result, then, is that philosophers using the words "thing" and "object" in subtly different ways reach different conclusions about what kinds of things or objects there are. But, of course, in using their central terms in different ways, they end up talking past one another.

A second way of responding to the neo-Carnapian is to point out that the argument for thinking that metaphysical disputes either have easy resolutions or are mere verbal disputes itself depends on substantive metaphysical premises.

So, for example, suppose I reason like this:

(1)   There are two prime numbers between four and eight.
(2)   In general, if there are two Fs between some x and some y, then there are Fs.
(3)   Therefore: There are (at least) two numbers.
(4)   Therefore: Numbers exist.

The neo-Carnapian declares that this argument is straightforwardly valid, and that the premises are clearly true; thus, the question "Are there numbers?" has an easy answer: "Yes, of *course* there are!" But not so fast. Just as the metaphysician asks "Are there really numbers?" so too they might ask "Are there really two prime numbers between four and eight?" Here the neo-Carnapian may roll their eyes and say, "Come on! Do you really mean to question our most basic and well-entrenched mathematical theories?" There is no reason why the metaphysician should be cowed by this response. After all, it is a substantive question whether the truth conditions for "there are two prime numbers between four and eight" require the existence of numbers.

A substantive question maybe; but must we think it is a substantive *metaphysical* question? Might it not be a substantive *linguistic* or *mathematical* question instead—one to be answered by examining the behavior of expressions like "There are" in English or in specifically mathematical contexts? Here I think that the answer is unequivocally and obviously "No". The reason is quite simple. Suppose linguistic or mathematical theory supplies you with an answer to a question of the form "Are there Fs?" Now consider the question whether the answer thus supplied is *really* correct—whether, in other words, there *really are* Fs. This is clearly a different question from the first one. Carnap, as we know, would have dismissed it as meaningless. But what sets the neo-Carnapians apart from Carnap himself is the refusal to join Carnap in this dismissal. But the question cannot be construed as just another linguistic or mathematical question either; for otherwise it would not be a different question. So it is a substantive question in metaphysics.

CONCLUSION

In this chapter, I have explored and responded to three different objections against metaphysics: the Humean complaint that metaphysicians rely illegitimately on intuitions, the Carnapian complaint that metaphysical questions are meaningless, and the neo-Carnapian complaint that metaphysical disputes either have easy or trivial answers or are nothing more than mere verbal disputes. Not everyone will be convinced by the responses I have given; nor, indeed, can *all* of

the responses be endorsed at once since different responses embody different visions about how exactly metaphysics is to be done. For my part, however, I do think that at least one of the responses is successful, that the typical debates in metaphysics are substantive, and that intuition is a perfectly legitimate source of evidence for premises in arguments for metaphysical conclusions. And so I encourage you, my reader, to join me in continuing forward in this exciting field of study!

FURTHER READING

The edition of Hume's *Enquiry* which I cited in the course of discussing objections against metaphysics is: *An Enquiry Concerning Human Understanding; [with] A Letter from a Gentleman to His Friend in Edinburgh; [and] An Abstract of a Treatise of Human Nature*, 2nd ed. (Indianapolis, IN: Hackett, 1993). A recent book that objects to traditional metaphysics in the same spirit as Hume did is James Ladyman and Don Ross, Every Thing Must Go: Metaphysics Naturalized, 1st ed. (New York: Oxford University Press, 2009).

In discussing Carnap's objection against metaphysics, I cited his "The Elimination of Metaphysics Through Logical Analysis of Language", *Erkenntnis* (1932): 60–81; but the more important source for the views I attributed to him is his "Empiricism, Semantics, and Ontology", in *Meaning and Necessity: A Study in Semantics and Modal Logic* (Chicago, IL: University of Chicago Press, 1956), 205–221. For discussion of Carnap's views and more detailed and sophisticated treatment of the Quinean response to Carnap, see Huw Price, "Metaphysics After Carnap : The Ghost Who Walks?", in *Metametaphysics*, 320–346, and Stephen Yablo "Does Ontology Rest on a Mistake?", *Aristotelian Society Supplementary Volume* 72 (1998): 229–261. The paper by Yablo also provides a *neo*-Carnapian perspective on metaphysics. For other neo-Carnapian perspectives, see the contributions by Eli Hirsch, Thomas Hofweber, and Amie Thomasson to the *Metametaphysics* volume, as well as Amie Thomasson's books, *Ordinary Objects* (Oxford: Oxford University Press, 2007) and *Ontology Made Easy* (Oxford: Oxford University Press, 2018).

For Quine's conception of metaphysics, I think that the best place to start is W. v. Quine, "Naturalism; Or, Living Within One's Means", *Dialectica* 49, no. 2–4 (1995): 251–263. But readers will also want to look at his "On What There Is", reprinted in Michael Rea, *Arguing about Metaphysics* (New York: Routledge, 2009), and Peter van Inwagen, "Being, Existence, and Ontological Commitment", in *Metametaphysics: New Essays on the Foundations of Ontology* (Oxford: Oxford University Press, 2009). Quine's article demonstrates Quine's approach to

answering metaphysical questions, and van Inwagen's article explains that approach.

I cited a passage from Reid's *Inquiry into the Human Mind* in defense of the evidential value of intuition. That can be found in *Thomas Reid's Inquiry and Essays*, edited by R. E. Beanblossom and K. Lehrer (Indianapolis, IN: Hackett, 1983), 168–169. For contemporary defenses of the evidential value of intuition, see George Bealer, "The Incoherence of Empiricism", *Aristotelian Society Supplementary Volume* 66 (1992): 99–138; Joel Pust, "Against Explanationist Skepticism Regarding Philosophical Intuitions", *Philosophical Studies* 106 (2001): 227–258. For a recent book-length defense of a broadly Reidian approach to a variety of skeptical considerations, see Michael Bergmann's *Radical Skepticism* (Oxford: Oxford University Press, 2021).

GLOSSARY

a priori Justified for a believer independently of sense perception or other modes of experience. (NB, a belief or proposition may count as *a priori* for a person even if her ability to *understand* it is somehow derived from experience.)

A–series A series of events whose members have properties like *being past*, *being present*, and *being future*.

A–theory of time A theory of time according to which time is both an A-series and a B-series.

agent causation Causation by the direct, immediate activity of an agent, neither by way of choices, beliefs, desires, or other events, nor as a matter of pure chance.

ameliorative definition A definition of a term (e.g., "woman") that does not aim to capture the dominant meaning of the terms, but is instead crafted to serve appropriate moral and political goals, such as the goal of eradicating certain kinds of oppression, doing away with problematic social hierarchies, and the like.

analytic All definitions of "analytic" are controversial; but the term usually applies to *sentences* that are commonly said to express conceptual truths, or to be true by definition, or to be such that their truth is grounded solely in the meanings of the terms involved.

B-series A series of events whose members are ordered by simultaneity and temporal priority (earlier-than and later-than) relations.

B-theory of time A theory of time according to which time, if it exists at all, is just a B-series.

class A collection of objects.

compatibilism The thesis that freedom is compatible with determinism.

constituent A rough synonym for "part" which, in philosophical contexts, is often used to signal that the things to which the term applies do not obey the axioms of standard logics of parthood.

constructed kinds Kinds whose members are such that their nature or existence depends directly or indirectly on human mental states. Commonly cited examples include *nation*, *dollar bill*, and *work of art*.

constructivism The view that, although there is *something* in the world that is neither a mind nor dependent upon minds, all of the kinds with which we human beings are acquainted are constructed kinds.

contingent Actual, or true, but not necessary. So a contingent truth is a proposition that is true but not necessarily true, and a contingent falsehood is one that is false but not necessarily false.

determinism The thesis that there is, at any given moment one and only one physically possible future.

empirical Empirical claims are claims that one can be justified in believing only on the basis of observation or experience

empiricist Someone who endorses *empiricism*, the view that all justification for non-analytic truths ultimately derives from experience rather than *a priori* reflection.

fundamental Fundamental things are the things out of which all others are made, or the things without which nothing else could exist.

frame of reference An object's frame of reference is a coordinate system that takes the object as a stationary origin point. Objects in motion relative to one another will therefore exist in different frames of reference.

gender essentialism The view that gender attributes are real essences. Gender essentialism, understood in this way, is usually developed and defended under the assumption that gender and

sex are binary, such that *man* and *woman* are the only two gender attributes, and *male* and *female* are the only two sexes.

hard determinism Hard determinism is the thesis that (i) determinism is true and (ii) incompatibilism is true. Hard determinism implies that no agent acts freely.

idealism The view that everything either is a mind or is among the contents of a mind.

immanent universal Something that is located in spacetime and has multiple instances. Properties and powers, like *humanity* or *negative charge* are often (controversially) taken to be immanent universals located exactly where their instances (e.g., individual humans or negatively charged particles) are located.

incompatibilism The thesis that freedom is not compatible with determinism.

indeterminism The thesis that determinism is false. Indeterminism might be true even if much in the universe is wholly determined by local causal factors.

intrinsic property A property that something can have regardless of what the rest of the world is like.

intuitions Intellectual experiences of the sheer obviousness or necessary truth of various claims.

irreflexive A relation R is irreflexive if, and only if, nothing bears R to itself. In other words: R is irreflexive if, and only if, aRb implies $a \neq b$.

libertarianism The thesis that (i) some agents act freely and (ii) freedom is incompatible with determinism.

material constitution Material constitution occurs whenever an object a and an object b share all of the same parts in common.

mere plurality Some objects that do not together compose anything but, for the sake of convenience, are collectively referred to by a term (like "football team") that looks like the name for an individual thing.

metaphysically possible Possible in the broadest sense of the term.

modal expressions Expressions that communicate possibility or necessity, e.g. *may, might, must.* Talk about *modality* is talk about what is possible or necessary in various different senses. The

different forms of modality include metaphysical, physical or nomological, logical, epistemic, and deontic.

modal properties Modal properties are properties that have modality built into them—properties like *being able to survive being squashed* or *being possibly made of wood*.

nominal essence An abstract idea, grounded in superficial similarities of things, that we use as a basis for classifying them.

nominalism Commonly, nominalism is identified either with the view that everything is particular (i.e., there are no universals) or with the view that everything is concrete (i.e., there are no abstract objects). This book takes the former position.

nomological Of or pertaining to the laws of nature.

ontology Commonly characterized as the study of *what there is*. It is the branch of study which focuses on *existence* claims of the sort studied by metaphysics, and on the logical consequences thereof.

persistence conditions Persistence conditions are modal properties of a certain kind, namely, facts about what a thing can and cannot survive.

physicalism The view that all properties and objects are reducible to the properties and objects posited by our best theories in physics.

proper part A proper part of a thing x is a part of x that is not an *improper part* of x. An improper part of x is any object that shares all of its parts with x.

real essence An intrinsic, complex attribute, or underlying structure, shared in common by the members of a kind, which serves to explain their superficial similarities and scientifically interesting features (e.g., behavior, natural development, overall appearance, and so on).

realism For any kind term "F" (e.g., "dog", "material object", "property"), realism about Fs is often characterized as the view that *there are Fs*, and that something's being an *F* does not depend on human mental activity—e.g., human beliefs, opinions, or concepts. Many philosophers, however—especially those who think that social kinds are both real and socially constructed—insist on a characterization of realism that does not make reference to mind independence.

reflexive A relation R is reflexive if, and only if, everything in its domain bears R to itself. In other words: R is reflexive if, and only if, for all x in the domain of R, xRx.

relata The relata of a relation are the things that stand in that relation.

relational property Relational properties are properties of an object that involve relations to other things.

social constructionism The view that at least some social kinds are constructed kinds. Views according to which a specific social kind, like gender, is socially constructed might be referred to by a term like "social constructionism about gender".

social metaphysics The sub-field of metaphysics that is concerned primarily with metaphysical questions about the objects, structures, and categories that are central to human social life, — e.g., gender, race, sex, disability, sexual orientation, the nature of groups, collective belief and action, and corporate personhood.

sparse theory of properties A theory of properties that places substantial restrictions on which predicates correspond to properties.

state of affairs A state of affairs is a circumstance, a situation, or a way things are.

substance Substances are property-bearers, subjects of predication. Sometimes the term is used simply to draw a contrast with attributes. More often, substances are construed as *independent* and *unified* things on which other things depend for their existence.

sum A sum of some objects or events, the xs, is an object or event that has all of the xs as parts and which is such that all of its parts overlap at least one of the xs.

symmetric A relation R is symmetric if, and only if, aRb implies bRa. The *resembles* relation is symmetric; the *is smarter than* relation is not.

thick particular A state of affairs which has only non-relational properties as constituents. (Non-relational properties are properties that do not involve relations, —e.g., properties like *having 2kg mass*.)

thin particular A state of affairs considered in abstraction from its properties.

transgender Individuals who identify as transgender are usually people who identify themselves as belonging to the gender opposite to the one that is commonly thought to accompany the sex they were assigned at birth. So, e.g., a trans woman is someone who was assigned "male" at birth but who identifies as

a woman, and a trans man is someone who was assigned "female" at birth but identifies as a man.

transitive A relation R is transitive if, and only if, aRb and bRc together imply aRc. The *is in the same room as* relation is transitive; the *loves* relation is not.

trope Abstract particular entities like *the particular whiteness of my shirt*, *the particular tomato-basil flavor of this spoonful of pizza sauce, Socrates's humanity*, etc. Tropes are typically referred to as "property-instances" because they are tokens of particular property types.

universal A universal is an abstract object which, in the paradigm case, either has or can have instances, or is either located or locatable in multiple places at once.

INDEX